



# STIC Search Report

## Biotech-Chem Library

STIC Database Tracking Number: 109338

TO: Manjunath N Rao  
Location: CM1/10A11&10D)1  
Art Unit: 1652  
December 10, 2003

Case Serial Number: 10/074527

From: P. Sheppard  
Location: CM1-1E03  
Phone: (703) 308-4499

sheppard@uspto.gov

### Search Notes

Examiner Search notes.

**THIS PAGE BLANK (USPTO)**

STIC-Biotech/ChemLib

109338

From: Rao, Manjunath N.  
Sent: Monday, December 01, 2003 10:07 AM  
T : STIC-Biotech/ChemLib  
Subject: Sequence search request for 10/074,527

From: Manjunath N. Rao  
Art Unit 1652, Room 10A11  
Mail Box in Room 10D 01  
Phone: 306-5681

CRFF

Date: 12-1-03

Please search the following as soon as possible for application with serial number

**10/074,527**

1. SEQ ID NO: 1 and 3 against all commercial nucleic acid databases, issued patents/published applications database and pending application database. Please provide a print of all results
2. SEQ ID NO: 2 against all commercial amino acid databases, issued patents/published applications database and pending application database. Please provide a print of all results.

If you have any questions please call me at the above phone number.

Thanks

Manjunath N. Rao, Ph.D.  
Biotechnology Patent Examiner  
Art Unit 1652, Room 10A11  
Mail Box in 10D01  
Crystal Mall 1, USPTO.

Point of Contact  
P. Sheppard  
telephone number: (703) 308-4499

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: \_\_\_\_\_  
Date Completed: 12/1/03  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:  
NA Sequences: \_\_\_\_\_  
AA Sequences: \_\_\_\_\_  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

**THIS PAGE BLANK (USPTO)**





**THIS PAGE BLANK (USPTO)**

PRIOR APPLICATION NUMBER: US 60/203,331  
PRIOR FILING DATE: 2000-05-11  
NUMBER OF SEQ ID NOS: 130  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 57  
LENGTH: 1746  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-292-896-57

Query Match 99.9%; Score 1744.4; DB 12; Length 1746;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1745; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	ATGTGGGGCGCACCGCGCGCGCGCTGCGCGCGGAACTGCGCGCGCGCGCGCGGAGCG	60
DB	1	ATGTGGGGCGCACCGCGCGCGCGCTGCGCGCGGAACTGCGCGCGCGCGCGGAGCG	60
QY	61	CTGTGGTGTCTTGGCGCTACTGCGCTTGGCGCGCGCGCGCGCGCGCGCGCGCG	120
DB	61	CTGTGGTGTCTTGGCGCTACTGCGCTTGGCGCGCGCGCGCGCGCGCGCGCGCG	120
QY	121	CGTGGGGCGGGCG	180
DB	121	CGTGGGGCGGGCG	180
QY	181	CGGAGCGCGGTATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	240
DB	181	CGGAGCGCGGTATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	240
QY	241	CGGAGCGCGGTATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	300
DB	241	CGGAGCGCGGTATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	300
QY	301	CACGAGATTAACATCTACCTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	360
DB	301	CACGAGATTAACATCTACCTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	360
QY	361	TGGAACCGCGTGTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	420
DB	361	TGGAACCGCGTGTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	420
QY	421	ATCATAGCATTTTATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	480
DB	421	ATCATAGCATTTTATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	480
QY	481	GAGACATCCCGGATATCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	540
DB	481	GAGACATCCCGGATATCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	540
QY	541	GAGCATTCCCGGATATCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	600
DB	541	GAGCATTCCCGGATATCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	600
QY	601	CGCGCCACACAG	660
DB	601	CGCGCCACACAG	660
QY	661	GCGGATGTTCTGACCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	720
DB	661	GCGGATGTTCTGACCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	720
QY	721	CTGCTGACAGAGATCCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	780
DB	721	CTGCTGACAGAGATCCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	780
QY	781	GACTGGACACCTTGAATACCTGGGAACTCCGGGAGAGAGAGAGAGAGAGAGAG	840
DB	781	GACTGGACACCTTGAATACCTGGGAACTCCGGGAGAGAGAGAGAGAGAGAGAG	840
QY	841	TGAGGCTGTGTTTACGTGACACACAGTCTTCTGAGAGAGAGAGAGAGAGAGAG	900
DB	841	TGAGGCTGTGTTTACGTGACACACAGTCTTCTGAGAGAGAGAGAGAGAGAGAG	900

## RESULT 5

US-09-925-299-219  
; Sequence 219, Application US/09925299  
; Patent No. US20020055627A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA102  
; CURRENT APPLICATION NUMBER: US/09/925,299  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05883

**THIS PAGE BLANK (USPTO)**

Db 121 WNPCKEKKYDYNLPTSVIIAFYNEAMSTLLRTVYVLETSPLDILLSEVILVDDYSDR 180  
Qy 181 EHLKERLANELSGLPKVLIRANKREGVLRARLLGASAARGDVLTFDCHCECHGWLEP 240  
Db 181 EHLKERLANELSGLPKVLIRANKREGVLRARLLGASAARGDVLTFDCHCECHGWLEP 240  
Qy 241 LLQRIHEESAVVCPVIDIDWNTFEYLGNSGSPQIGGFDWRLVFTWHTVPERIRMQS 300  
Db 241 LLQRIHEESAVVCPVIDIDWNTFEYLGNSGSPQIGGFDWRLVFTWHTVPERIRMQS 300  
Qy 301 PVDVIRSPMTAGGLFAVSKYFEYLGNSGSPQIGGFDWRLVFTWHTVPERIRMQS 360  
Db 301 PVDVIRSPMTAGGLFAVSKYFEYLGNSGSPQIGGFDWRLVFTWHTVPERIRMQS 360  
Qy 361 VGHVFPKQAPYSRNKALANSVRAAEVMMDEFKELYHHRNPRARLEPFGDVTERRKQRLDKL 420  
Db 361 VGHVFPKQAPYSRNKALANSVRAAEVMMDEFKELYHHRNPRARLEPFGDVTERRKQRLDKL 420  
Qy 421 QCKDFKWFLETVPELHVPEDRPGFGMLQNKGLTDYCFDYNPPDENQIVGHQVILYLCH 480  
Db 421 QCKDFKWFLETVPELHVPEDRPGFGMLQNKGLTDYCFDYNPPDENQIVGHQVILYLCH 480  
Qy 481 GMGQNOFFEYTSQKEIRYNTHQPEGCIATVEAGMDTLIMHLCEETAPENOKFILOEDGSLF 540  
Db 481 GMGQNOFFEYTSQKEIRYNTHQPEGCIATVEAGMDTLIMHLCEETAPENOKFILOEDGSLF 540  
Qy 541 HEQSKKCVQAARKESDSSFVPLLRDCTNSDHQKWFKKERML 581  
Db 541 HEQSKKCVQAARKESDSSFVPLLRDCTNSDHQKWFKKERML 581

## RESULT 2

US-10-292-896-58

; Sequence 58, Application US/10292896

; Publication No. US20030186850A1

; GENERAL INFORMATION:

; APPLICANT: HASSAN, Helle

; APPLICANT: REIS, Celso A.

; APPLICANT: BENNETT, Eric P.

; APPLICANT: CLAUSEN, Henrik

; TITLE OF INVENTION: METHODS TO IDENTIFY AGENTS MODULATING FUNCTIONS OF POLYPEPTIDE GA

; TITLE OF INVENTION: TRANSPERASES, PHARMACEUTICAL COMPOSITIONS COMPRISING SUCH AGENTS

; TITLE OF INVENTION: SUCH AGENTS FOR PREPARING MEDICAMENTS

; FILE REFERENCE: 4305/1H154-US3

; CURRENT APPLICATION NUMBER: US/10/292,896

; CURRENT FILING DATE: 2003-03-17

; PRIOR APPLICATION NUMBER: US 60/425,204

; PRIOR FILING DATE: 2002-11-08

; PRIOR APPLICATION NUMBER: PCT/DK01/00328

; PRIOR FILING DATE: 2001-05-10

; PRIOR APPLICATION NUMBER: US 60/203,331

; PRIOR FILING DATE: 2000-05-11

; NUMBER OF SEQ ID NOS: 130

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 58

; LENGTH: 581

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-292-896-58

Query Match 98.5%; Score 3076; DB 12; Length 581;

Best Local Similarity 98.5%; Pred. No. 2.4e-278;

Matches 572; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 1 MWGTRARRCPRELRRGREALLVALLLAGLGSVLRARQAGAGAAEPGPPRTPRGR 60

Db 1 MWGTRARRCPRELRRGREALLVALLLAGLGSVLRARQAGAGAAEPGPPRTPRGR 60

Qy 61 REPVMPRPVPANALGARGAVRLQOGEELRLQOESVRLHQINIVLSRISLHRLRPER 120

Db 61 REPVMPRPVPANALGARGAVRLQOGEELRLQOESVRLHQINIVLSRISLHRLRPER 120

Qy 121 WNPCKEKKYDYNLPTSVIIAFYNEAMSTLLRTVYVLETSPLDILLSEVILVDDYSDR 180

## RESULT 3

US-10-074-527-8

; Sequence 8, Application US/10074527

; Publication No. US20020142426A1

; GENERAL INFORMATION:

; APPLICANT: Olandt, Peter J.

; APPLICANT: Meyers, Rachel E.

; APPLICANT: Galvin, Katherine A.

; APPLICANT: Millennium Pharmaceuticals Inc.

; TITLE OF INVENTION: 33945, A Human Glycosyltransferase and

; TITLE OF INVENTION: Uses Therefor

; FILE REFERENCE: MPI2001-018PIRCPI (M)

; CURRENT APPLICATION NUMBER: US/10/074,527

; CURRENT FILING DATE: 2002-02-12

; PRIOR APPLICATION NUMBER: 60/269202

; PRIOR FILING DATE: 2001-02-15

; NUMBER OF SEQ ID NOS: 9

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 8

; LENGTH: 578

; TYPE: PRT

; ORGANISM: Mus musculus

US-10-074-527-8

Query Match

Best Local Similarity 57.9%; Pred. No. 2e-147;

Matches 330; Conservative 78; Mismatches 144; Indels 18; Gaps 8;

Qy 23 VILLALLAGL-----GSVLRARQAGAGAAEPGPPRTPRGRRE-----PVMPPRPVPA 72

Db 13 LLLALLTAYITLVFVSSTLYASPGAG-GARELGPRLPDLDTRFEEDLSQPLYTKPPADS 71

Qy 73 NALGARGAVRLQOGEELRLQOESVRLHQINIVLSRISLHRLRPERWNPCKEKKYDY 132

Db 72 HALGEWGRASKQLNEGELKQOEEELIERYALINIVLSRISLHRIEDKRYTECKAKPHY 131

Qy 133 DNLPRTSVIIAFYNEAMSTLLRTVYVLETSPLDILLSEVILVDDYSDRHLKERLANELS 192

Db 132 RSLPTTSVIIAFYNEAMSTLLRTVYVLETSPLDILLSEVILVDDYSDRILYKAQLETYS 191

Qy 193 GLPKVRLIRANKREGVLRARLLGASAARGDVLTFDCHCECHGWLEPFLQRIHEESAV 252

**THIS PAGE BLANK (USPTO)**

XX The invention relates to an isolated human drug metabolising enzyme (DME)  
CC and its nucleotide. DME is useful for diagnosing, treating or preventing  
CC disorders associated with aberrant expression of DME, where the disorders  
CC are selected from autoimmune/inflammatory disorder such as acquired  
CC immunodeficiency syndrome (AIDS), asthma, atherosclerosis, psoriasis,  
CC uveitis; a cell proliferative disorder such as arteriosclerosis,  
CC cirrhosis, hepatitis, and cancer; a neurological disorder such as  
CC Alzheimer's disease, Huntington's disease, dementia, Parkinson's disease;  
CC a developmental disorder such as renal tubular acidosis, epilepsy,  
CC anaemia; an endocrine disorder such as adenoma, thrombosis and  
CC infections; an eye disorder such as conjunctivitis, glaucoma, cataract;  
CC metabolic disorder such as cystic fibrosis, diabetes and goitre; a  
CC gastrointestinal disorder such as anorexia, peptic ulcer; and liver  
CC disorders. DME is useful in a number of drug screening techniques and to  
CC analyse the proteome of a tissue or cell type. The invention is useful  
CC for creating knock-in humanised animals or transgenic animals to model  
CC human diseases, in somatic or germline gene therapy, to generate a  
CC transcript image of a tissue or cell type, for detecting differences in  
CC the chromosomal location due to translocation, inversion, etc. among  
CC normal, carrier or affected individuals, and as hybridisation probes for  
CC mapping naturally occurring genomic sequences. The present sequence is  
CC human DME-4.  
XX

SQ Sequence 581 AA;

Query Match 100.0%; Score 3124; DB 23; Length 581;  
Best Local Similarity 100.0%; Pred. No. 8.6e-292;  
Matches 581; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MNGTARRCCPELRGREALVLLALLAGLSVLAQAGAGAAEPPGPPRPRGR 60  
Db 1 MNGTARRCCPELRGREALVLLALLAGLSVLAQAGAGAAEPPGPPRPRGR 60  
Qy 61 REPVMPRPVPANALGARGEAVRLQOGEELRLOESVRLHQINLYSDRLSLHRLPER 120  
Db 61 REPVMPRPVPANALGARGEAVRLQOGEELRLOESVRLHQINLYSDRLSLHRLPER 120  
Qy 121 WNPLCKEKYDYNLPRTSVIIAFYNEAWSTLLRTVYSLVLETSPIILLEEVILVDDYSDR 180  
Db 121 WNPLCKEKYDYNLPRTSVIIAFYNEAWSTLLRTVYSLVLETSPIILLEEVILVDDYSDR 180  
Qy 181 EHLKERLANELSGLPKVLIRANKREGLVRALLGASAAARGDVLFTLDCHEGWLPEP 240  
Db 181 EHLKERLANELSGLPKVLIRANKREGLVRALLGASAAARGDVLFTLDCHEGWLPEP 240  
Qy 241 LLQRIHEESAVVCPVIDVDMNTFEYLGNSGEPOIGGDFMLVFTWHTVPERIRMQS 300  
Db 241 LLQRIHEESAVVCPVIDVDMNTFEYLGNSGEPOIGGDFMLVFTWHTVPERIRMQS 300  
Qy 301 PVDVIRSPMTAGGLFAVSKKYEYLGSDYDTGMEVNGGENLEFSFRIWQCGGVLETHPCSH 360  
Db 301 PVDVIRSPMTAGGLFAVSKKYEYLGSDYDTGMEVNGGENLEFSFRIWQCGGVLETHPCSH 360  
Qy 361 VGHVFPKQAPYGRNKKALANSVRAAEVWDMDFKELYHNRNPRARLPFGDVTERRKQRLK 420  
Db 361 VGHVFPKQAPYGRNKKALANSVRAAEVWDMDFKELYHNRNPRARLPFGDVTERRKQRLK 420  
Qy 421 QCKDFKWLFTVYVPELVHPEDRPGFGLMQLNKGLTDYCFDYNPPDENQIVGHQVILYLCH 480  
Db 421 QCKDFKWLFTVYVPELVHPEDRPGFGLMQLNKGLTDYCFDYNPPDENQIVGHQVILYLCH 480  
Qy 481 GNGQNOFFEYTSQKEIRYNTHOPEGCIAVEAGMDTLIMHLCBETAPENKFILQEDGSILF 540  
Db 481 GNGQNOFFEYTSQKEIRYNTHOPEGCIAVEAGMDTLIMHLCBETAPENKFILQEDGSILF 540  
Qy 541 HQSKKVCVQAARKESSDSFVPLLRDCTNSDHOQWFFKERM 581  
Db 541 HQSKKVCVQAARKESSDSFVPLLRDCTNSDHOQWFFKERM 581

RESULT 3  
ABG04767

ID ABG04767 standard; Protein; 355 AA.  
XX  
AC ABG04767;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
XX Novel human diagnostic protein #4758.  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder.  
XX Homo sapiens.  
OS  
XX WO200175067-A2.  
XX  
XX 11-OCT-2001.  
XX  
XX 30-MAR-2001; 2001WO-US08631.  
XX  
XX 31-MAR-2000; 2000US-0540217.  
XX  
XX 23-AUG-2000; 2000US-0649167.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Drmanac RT, Liu C, Tang YT;  
XX  
XX WPI; 2001-639362/73.  
XX  
XX N-PSDB; AAS68954.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits and to assess  
XX biodiversity.  
XX  
XX Claim 20; SEQ ID No 35126; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX and gene mapping, and in recombinant production of (II). The  
XX polynucleotides are also used in diagnostics as expressed sequence tags  
XX for identifying expressed genes. (I) is useful in gene therapy techniques  
XX to restore normal activity of (II) or to treat disease states involving  
XX (II). (II) is useful for generating antibodies against it, detecting or  
XX quantitating a polypeptide in tissue, as molecular weight markers and as  
XX a food supplement. (II) and its binding partners are useful in medical  
XX imaging of sites expressing (II). (I) and (II) are useful for treating  
XX disorders involving aberrant protein expression or biological activity.  
XX The polypeptide and polynucleotide sequences have applications in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX and to produce other types of data and products dependent on DNA and  
XX amino acid sequences. ABG00010-ABG30377 represent novel human  
XX diagnostic amino acid sequences of the invention.  
XX Note: The sequence data for this patent did not appear in the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 355 AA;

Query Match 57.2%; Score 1788; DB 22; Length 355;  
Best Local Similarity 98.5%; Pred. No. 2.5e-163;  
Matches 335; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
Qy 65 MPRPPVPANALGARGEAVRLQOGEELRLOESVRLHQINLYSDRLSLHRLPERVWNP 124  
Db 1 MPRPPVPANALGARGEAVRLQOGEELRLOESVRLHQINLYSDRLSLHRLPERVWNP 60  
Qy 125 CKEKKYDYNLPRTSVIIAFYNEAWSTLLRTVYSLVLETSPIILLEEVILVDDYSDRHLK 184  
Db 61 CKEKKYDYNLPRTSVIIAFYNEAWSTLLRTVYSLVLETSPIILLEEVILVDDYSDRHLK 120  
Qy 185 ERLANELSGLPKVLIRANKREGLVRALLGASAAARGDVLFTLDCHEGWLPELQ 244

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6

Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 7, 2003, 12:45:48 ; Search time 6368.95 seconds  
(without alignments)  
11215.070 Million cell updates/sec

Title: US-10-074-527-3

Perfect score: 1746

Sequence: 1 atgtggtggggcgacggcgcg.....tcaagagcgcatgttatga 1746

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*

1: gb\_ba:\*

2: gb\_btg:\*

3: gb\_in:\*

4: gb\_om:\*

5: gb\_ov:\*

6: gb\_pat:\*

7: gb\_ph:\*

8: gb\_pl:\*

9: gb\_pr:\*

10: gb\_ro:\*

11: gb\_sts:\*

12: gb\_sy:\*

13: gb\_un:\*

14: gb\_vi:\*

15: em\_ba:\*

16: em\_fun:\*

17: em\_hum:\*

18: em\_in:\*

19: em\_mu:\*

20: em\_om:\*

21: em\_or:\*

22: em\_ov:\*

23: em\_pat:\*

24: em\_ph:\*

25: em\_pl:\*

26: em\_ro:\*

27: em\_sts:\*

28: em\_un:\*

29: em\_vi:\*

30: em\_htg\_hum:\*

31: em\_htg\_inv:\*

32: em\_htg\_other:\*

33: em\_htg\_mus:\*

34: em\_htg\_pln:\*

35: em\_htg\_rod:\*

36: em\_htg\_mam:\*

37: em\_htg\_vrt:\*

38: em\_sy:\*

39: em\_htgo\_hum:\*

40: em\_htgo\_mus:\*

41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score on the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1746	100.0	2745	6	AX477705	AX477705 Sequence
2	1744.4	99.9	1746	9	AB078146	AB078146 Homo sapi
3	1744.4	99.9	1746	9	HS132365	HS132365 Homo sapi
4	1319	75.5	2343	9	AK024865	AK024865 Homo sapi
5	851.4	48.8	63632	2	AC132018.4	Continuation (5 of
c	851.4	48.8	110000	2	AC098559.2	Continuation (3 of
6	851.4	48.8	288118	2	AC129139	AC129139 Rattus no
7	851.4	48.8	228368	2	AC121727	AC121727 Rattus no
8	842.6	48.2	94616	2	AC098559.7	Continuation (8 of
c	841	48.2	288118	2	AC129139	AC129139 Rattus no
9	830	47.5	1928	9	BC013945	BC013945 Homo sapi
10	812.6	46.5	110000	2	AC099237.3	Continuation (4 of
11	812.6	46.5	166547	2	AC134060	AC134060 Rattus no
c	805.4	46.1	166547	2	AC134060	AC134060 Rattus no
12	803.8	46.0	110000	2	AC098559.1	Continuation (2 of
13	795.4	45.6	110000	2	AC095991.0	AC095991 Rattus no
c	795.4	45.6	110000	2	AC112324.0	AC112324 Rattus no
14	795.4	45.6	110000	2	AC112324.0	Continuation (3 of
15	795.4	45.6	299134	2	AC128783	AC128783 Rattus no
16	792.2	45.4	110000	2	AC105836.1	Continuation (2 of
c	792.2	45.4	110000	2	AC105836.1	AC105836 Rattus no
17	522.4	29.9	226623	2	AC097343	AC097343 Rattus no
18	520.8	29.8	1737	10	MMU73819	U73819 Mus musculu
19	510.6	29.2	110000	2	AC101864	AC101864 Mus muscu
20	499.4	28.6	5369	9	BC036390	BC036390 Rattus no
21	499.4	28.6	73481	9	BC036390	BC036390 Homo sapi
22	496.2	28.4	1737	6	AR236647	AR236647 Homo sapi
23	496.2	28.4	1737	9	HSY08564	HSY08564 Sequence
c	493.6	28.3	165175	2	AC083808	Y08564 Homo sapien
24	419	24.0	554	6	AX197983	AC083808 Homo sapi
25	418.4	24.0	544	6	AX208574	AX197983 Sequence
26	400.8	23.0	652	6	AX197725	AX208574 Sequence
27	400.2	22.9	632	6	AX208332	AX197725 Sequence
28	400.2	22.9	632	6	AX209583	AX208332 Sequence
29	374	21.4	149930	9	AL136084	AX209583 Sequence
30	363	20.8	483	6	AX371123	AL136084 Human DNA
31	300.8	17.2	2384	3	AK112431	AX371123 Sequence
32	285.6	16.4	1920	6	AR153422	AK112431 Clona int
33	285.6	16.4	2575	6	AX376296	AR153422 Sequence
34	285.6	16.4	2575	6	AX697278	AX376296 Sequence
35	285.6	16.4	3357	9	AY035399	AX697278 Sequence
36	285.6	16.4	3580	9	HSN803261	AY035399 Homo sapi
37	284	16.3	3332	6	AR155987	AL831925 Homo sapi
38	284	16.3	3528	9	BC014789	AR155987 Sequence
39						BC014789 Homo sapi

## ALIGNMENTS

RESULT 1

AX477705

LOCUS AX477705

DEFINITION Sequence 17 from Patent WO0246426.

ACCSSION AX477705

VERSION AX477705.1

KEYWORDS GI:22216860

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1

REFERENCE

AUTHORS

Sanjanwala, M.M., Yao, M.G., Au-Young, J., Baughn, M.R., Arvizu, C.,

Ring, H.Z., Lee, E.A., Ding, L., Hafalia, A.J., Tang, Y.T., Yue, H.,

Tribouley, C.M., Lu, D.A., Lal, P.G., Warren, B.A., Yang, J.,

linear PAT 12-AUG-2002

**THIS PAGE BLANK (USPTO)**

Walia, N.K., Nguyen, D.B., Gandhi, A.R. and Ison, C.H.

Drug metabolizing enzymes  
Patent: WO 0246426-A 17 13-JUN-2002;  
Incyte Genomics, Inc. (US)

FEATURES	Location/Qualifiers
SOURCE	1. 2745

```
1. 2743
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
```

```

BASE COUNT
704 a . 612 c 737 q 692 f
/Hoste="Incyte ID No: 2860635CBI

```

Query Match	100.0%	Score 1746	DB 6	Length 2745

Best Local Similarity	100.0%	Pred. No. 2e-266;	
Matches 1746;	Conservative	0;	Mismatches 0;
		Indels	0;
		Gaps	0;

Qy 1 ATGTGGGGCGGCACGGCGCGCGGGCGCTGCCCGCGGAACTGCGCGCGGGCCCGGAGGCG 60  
|||  
Db 1 ATGTGGGGCGGCACGGCGCGCGGGCGCTGCCCGCGGAACTGCGCGCGGGCCCGGAGGCG 60  
|||

QY 61 CTGTTGGTGCTCTGGCGGCTACTGGCGCTTGGCCGGGCTGGGCTCGGTCGCGGCGGAG 120  
 |||||  
 Db 61 CTGTTGGTGCTCTCTGGCGGCTACTGGCGCTTGGCCGGGCTGGGCTCGGTCGCGGCGGAG 120  
 |||||

QY 121 CGTGGGGCCGGGGCCGGGGCTGCGAGCCCGGACCCCGCGGACCCCGCGGCCCGGGCGG 180

Db 121 CGTGGGGCCGGGGCCGGGGCTGCGAGCCCGGACCCCGCGGACCCCGCGGCCCGGGCGG 180

181	CGCGAGCCGGTCTATCGCGCGCGCGCGCGTCCGGCCGAA	CGCGCTGGGCGCGCGCGGGCGAG	240
Qy			
181	CGCGAGCCGGTCTATCGCGCGCGCGCGTCCGGCCGAA	CGCGCTGGGCGCGCGCGGGCGAG	240
Db			

Oy  
db

CGCTGGGCTGCAGCTGCAGGCCGACCTCCGGCTCAGGAGGAGCGTCGGGTG 300  
|||||  
  
CGCTGGGCTGCAGCTGCAGGCCGACCTCCGGCTCAGGAGGAGCGTCGGGTG 300

QY		301	CACCAGATTAACTTACCTCAGGACCGCATCTCACTGCACCGCGGCTGTCCCGAGGGC	360
db		301	CACCAGATTAACTTACCTCAGGACCGCATCTCACTGCACCGCGGCTGTCCCGAGGGC	360

361 TGGAACCCGGTGTGCAAGAGAGAGAAATATGATTATGATAATTTGCCCAGGACATCTGTT 420  
 361 TGGAACCCGGTGTGCAAGAGAGAGAAATATGATTATGATAATTTGCCCAGGACATCTGTT 420

421 ATCATAGCATTTTATATGAAGCGCTGGTCAACTCTCCTTCGGACAGTTTACAGTGTCTCT 480  
|||||  
421 ATCATAGCATTTTATATGAAGCGCTGGTCAACTCTCCTTCGGACAGTTTACAGTGTCTCT 480  
|||||

481 GAGACATCCCGGATATCTTGTAGAGAGTGATCTTTGTAGATGACTACAGTGATAGA 540  
|||||  
481 GAGACATCCCGGATATCTTGTAGAGAGTGATCTTTGTAGATGACTACAGTGATAGA 540  
|||||

541 GAGCAGCTGAAGGAGCGCTGGCCAATGAGCTTTGGGACTGCCCAAGGTGGCGCTGATC 600  
|||||  
541 GAGCAGCTGAAGGAGCGCTGGCCAATGAGCTTTGGGACTGCCCAAGGTGGCGCTGATC 600  
|||||

601 CGCGCCAAAGAGAGAGAGGCCCTGGTGCAGAGCCCGGCTGCTGGGGGGCGCTCTGCGCGCAGG 660

601 CGCGCCAAAGAGAGAGAGGCCCTGGTGCAGAGCCCGGCTGCTGGGGGGCGCTCTGCGCGCAGG 660

661 GCGCATGTTTGACCTTCTTGGACTCTCACTGTGAGTGCACGAAGGCTGGAGCCG 720  
 |||||  
 661 GCGCATGTTTGACCTTCTTGGACTCTCACTGTGAGTGCACGAAGGCTGGAGCCG 720  
 |||||

721 CTGTCGAGAGATCCATCAAGAGGAGTCGGCAGTGTGTGCCGGTGATTGATGTCATC 780  
|||||  
721 CTGTCGAGAGATCCATCAAGAGGAGTCGGCAGTGTGTGCCGGTGATTGATGTCATC 780  
|||||

Qy 781 GACTGGAAACACCTTCGAATACCTGGGGAACCTCGGGGAGCCCCAGATCGGGCGGTTTCGAC 840

Db 781 GACTGGAAACACCTTCGAATACCTGGGGAACCTCGGGGAGCCCCAGATCGGGCGGTTTCGAC 840

Qy	841	GGAGGCTGTGTTCCTGGCCACACAGTTCTCTGAGAGGAGAGGATACGGATGCAATCC	900
Db	841	GGAGGCTGTGTTCCTGGCCACACAGTTCTCTGAGAGGAGAGGATACGGATGCAATCC	900
Qy	901	CCGTCGATCTCATCAGGTCCTCAACAAATGGCTGGTGGGCTGTTTGTGTGAGTAAGAAA	960
Db	901	CCGTCGATCTCATCAGGTCCTCAACAAATGGCTGGTGGGCTGTTTGTGTGAGTAAGAAA	960
Qy	961	TATTTTCAATATCTGGGGTCTTATGATACAGGAATGGAAGTTTGGGAGGAGAAAACCTC	1020
Db	961	TATTTTCAATATCTGGGGTCTTATGATACAGGAATGGAAGTTTGGGAGGAGAAAACCTC	1020
Qy	1021	GAATTTTCCCTTAGGATCTGGCAGTGTGGTGGGTTCTGGAACACACCCATGTTCCCAT	1080
Db	1021	GAATTTTCCCTTAGGATCTGGCAGTGTGGTGGGTTCTGGAACACACCCATGTTCCCAT	1080
Qy	1081	GTTGGCCATGTTTTCCCAAGCAGCTCCCTACTCTCCGCCAACAAAGGCTCTGGGCCAACAGT	1140
Db	1081	GTTGGCCATGTTTTCCCAAGCAGCTCCCTACTCTCCGCCAACAAAGGCTCTGGGCCAACAGT	1140
Qy	1141	GTTCTGTGCAGCTGAAGTATGGATGGATGAATTTAAAGAGGCTCTACTACCATCGCAACCCC	1200
Db	1141	GTTCTGTGCAGCTGAAGTATGGATGGATGAATTTAAAGAGCTCTACTACCATCGCAACCCC	1200
Qy	1201	CGTGCCCGCTGGAACTTTTGGGGATGTGACAGAGAGGAAGCAGCTCCGGGCAAGCTC	1260
Db	1201	CGTGCCCGCTGGAACTTTTGGGGATGTGACAGAGAGGAAGCAGCTCCGGGCAAGCTC	1260
Qy	1261	CAGGTAAAGACATTCAGTGGTCTTTGGAGACTGTGTATCCAGAACTGCATGTGCCTGAG	1320
Db	1261	CAGGTAAAGACATTCAGTGGTCTTTGGAGACTGTGTATCCAGAACTGCATGTGCCTGAG	1320
Qy	1321	GACAGGCTGGCTTCTTCGGGATGCTCCAGAACAAAGGACTAACAGACTACTGCTTTGAC	1380
Db	1321	GACAGGCTGGCTTCTTCGGGATGCTCCAGAACAAAGGACTAACAGACTACTGCTTTGAC	1380
Qy	1381	TATAACCTCCCGATGAAACACAGATGTGGGACACCGGTCATCTGTACCTCTGTCAT	1440
Db	1381	TATAACCTCCCGATGAAACACAGATGTGGGACACCGGTCATCTGTACCTCTGTCAT	1440
Qy	1441	GGATGGGCCAGAACTCAGTTTTTCGAGTACAGTCCCGAAGAAATAGCCTATAACACC	1500
Db	1441	GGATGGGCCAGAACTCAGTTTTTCGAGTACAGTCCCGAAGAAATAGCCTATAACACC	1500
Qy	1501	CACCAAGCTGAGGGCTGCATTCGTGTGAAGCAGGAATGGATACCCCTTATCATGCATCTC	1560
Db	1501	CACCAAGCTGAGGGCTGCATTCGTGTGAAGCAGGAATGGATACCCCTTATCATGCATCTC	1560
Qy	1561	TGGGAAGAACTGCCCCAGAGAACTCAGAAATGTCCTTGCAGGAGGATCGATCTTTATTT	1620
Db	1561	TGGGAAGAACTGCCCCAGAGAACTCAGAAATGTCCTTGCAGGAGGATCGATCTTTATTT	1620
Qy	1621	CACGAACGTCCAGAAATGTCGTCCAGGCTCGGAGGAGGAGTCCAGTGCAGTTTCGTT	1680
Db	1621	CACGAACGTCCAGAAATGTCGTCCAGGCTCGGAGGAGGAGTCCAGTGCAGTTTCGTT	1680
Qy	1681	CCACTCTTACGAGACTGCACCAACTCGGATCATCAGAAATGTTCTTCAAAGAGGCGATG	1740
Db	1681	CCACTCTTACGAGACTGCACCAACTCGGATCATCAGAAATGTTCTTCAAAGAGGCGATG	1740
Qy	1741	TTATGA	1746
Db	1741	TTATGA	1746

RESULT 2  
AB078146

**LOCUS**  
**DEFINITION**

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

ACCESSION  
VERSION

AB078146 Homo sapiens GALNT12 mRNA for UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 12, complete cds. 1746 bp mRNA linear PRI 04-SEP-2002

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 7, 2003, 12:45:48 ; Search time 10396.1 Seconds  
(without alignments)  
11215.070 Million cell updates/sec

Title: US-10-074-527-1  
Perfect score: 2850  
Sequence: 1 ccgctcggtaccactataa.....aaaaaaaaaaaaaaaaaaaaa 2850

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*

1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*  
29: em.vi.\*  
30: em.htg.hum.\*  
31: em.htg.inv.\*  
32: em.htg.other.\*  
33: em.htg.mus.\*  
34: em.htg.pln.\*  
35: em.htg.rod.\*  
36: em.htg.man.\*  
37: em.htg.vrt.\*  
38: em.sv.\*  
39: em.htgo.hum.\*  
40: em.htgo.mus.\*  
41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to  
and is derived by analysis of the score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2738.2	96.1	2745	6	AX477705	Sequence AX477705
2	2341	82.1	2343	9	AK024865	Sequence AK024865
3	1838.6	64.5	1928	9	BC013945	Sequence BC013945
4	1744.4	61.2	1746	9	AB078146	Sequence AB078146
5	1744.4	61.2	1746	9	HS132365	Sequence HSA132365
6	1127	39.5	14930	9	AL136084	Sequence AL136084
7	1053.2	37.0	1808	9	AB048801	Sequence AB048801
8	1039	36.5	52737	2	AC007800	Sequence AC007800
9	921.6	32.3	63632	2	AC132018	Sequence AC132018
10	921.6	32.3	110000	2	AC098559	Sequence AC098559
11	911.4	32.0	288118	2	AC129139	Sequence AC129139
12	907.6	31.8	228368	2	AC121727	Sequence AC121727
13	906	31.8	288118	2	AC129139	Sequence AC129139
14	887.4	31.1	110000	2	AC099237	Sequence AC099237
15	887.4	31.1	166547	2	AC134060	Sequence AC134060
16	873.4	30.6	110000	2	AC098559	Sequence AC098559
17	870.4	30.5	166547	2	AC134060	Sequence AC134060
18	865	30.4	110000	2	AC095991	Sequence AC095991
19	865	30.4	110000	2	AC112324	Sequence AC112324
20	865	30.4	110000	2	AC115635	Sequence AC115635
21	865	30.4	299134	2	AC128783	Sequence AC128783
22	853	29.9	110000	2	AC105836	Sequence AC105836
23	844.8	29.6	94616	2	AC098559	Sequence AC098559
24	550	19.3	110000	2	AC099237	Sequence AC099237
25	527	18.5	554	6	AX197983	Sequence AX197983
26	526.4	18.5	544	6	AX208574	Sequence AX208574
27	526.2	18.5	328901	2	AX208574	Sequence AX208574
28	522.4	18.3	226623	2	AC097343	Sequence AC097343
29	520.8	18.3	1737	10	MMU73819	Sequence MMU73819
30	520.8	18.3	234109	2	AC101864	Sequence AC101864
31	499.4	17.5	5369	9	BC036390	Sequence BC036390
32	499.4	17.5	73481	9	AC025034	Sequence AC025034
33	497	17.4	652	6	AX197725	Sequence AX197725
34	496.4	17.4	632	6	AX208332	Sequence AX208332
35	496.4	17.4	632	6	AX209583	Sequence AX209583
36	496.2	17.4	1737	6	AR236647	Sequence AR236647
37	496.2	17.4	1737	9	HSV08564	Sequence HSV08564
38	493.6	17.3	165175	2	AC083808	Sequence AC083808
39	409	14.4	421	6	AX330041	Sequence AX330041
40	409	14.4	421	6	AX335657	Sequence AX335657
41	363	12.7	483	6	AX371123	Sequence AX371123
42	300.8	10.6	2384	3	AK112431	Sequence AK112431
43	285.6	10.0	1920	6	AR153422	Sequence AR153422
44	285.6	10.0	2575	6	AX376296	Sequence AX376296
45	285.6	10.0	2575	6	AX697278	Sequence AX697278

# ALIGNMENTS

RESULT 1  
AX477705  
LOCUS AX477705  
DEFINITION Sequence 17 from Patent W00246426.  
ACCESSION AX477705  
VERSION AX477705.1  
KEYWORDS GI:22216860  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

2745 bp DNA linear PAT 12-AUG-2002

REFERENCE  
AUTHORS  
Sanjanwala, M.M., Yao, M.G., Au-Young, J., Baughn, M.R., Arvizu, C.,  
King, H.Z., Lee, E.A., Ding, L., Hafalia, A.J., Tang, Y.T., Yue, H.,  
Tribouley, C.M., Lu, D.A., Lal, P.G., Warren, B.A., Yang, J.,

**THIS PAGE BLANK (USPTO)**



**THIS PAGE BLANK (USPTO)**



```
Db 1921 GCTTGTACTATTTTGAAGACTTTTAAATGTTCCAAATACCCCTATTTCCTCAAGGTA 1980
Qy 2061 ATCGTAAGATGTAACCCCTTGGTATTAGAAAATTAACACCTTTAATAATTTTCTATCA 2120
Db 1981 ATCGTAAGATGTAACCCCTTGGTATTAGAAAATTAACACCTTTAATAATTTTCTATCA 2040
Qy 2121 ABAWAWAATTTTACAGTCGTCCTTTACTCTCATATTAGCAAAAAGATAAGATTTTAT 2180
Db 2041 AGATGTATATTTTACAGTCGTCCTTTACTCTCATATTAGCAAAAAGATAAGATTTTAT 2100
Qy 2181 TTTGTATTACAAAGATTTCCAGGTACGAAGATATCTGATGGTGGAAATCAGTTCA 2240
Db 2101 TTTGTATTACAAAGATTTCCAGGTACGAAGATATCTGATGGTGGAAATCAGTTCA 2160
Qy 2241 AGCAACGTAATTTGCAATTAACCTGATTAACCTGCGGGGTTAAAGTTTTCACGTA 2300
Db 2161 AGCAACGTAATTTGCAATTAACCTGATTAACCTGCGGGGTTAAAGTTTTCACGTA 2220
Qy 2301 TAGAGAGACTGTCACCTAGCAATTTGATTTATTTATTCAGGTCAATGAGATCTTCTAGA 2360
Db 2221 TAGAGAGACTGTCACCTAGCAATTTGATTTATTTATTCAGGTCAATGAGATCTTCTAGA 2280
Qy 2361 TGTATTTTAAAAAGAAATGCTTTTGGTTATGTTGTTGCTACACAGTTAAAGTTCATAAT 2420
Db 2281 TGTATTTTAAAAAGAAATGCTTTTGGTTATGTTGTTGCTACACAGTTAAAGTTCATAAT 2340
Qy 2421 GTTCATGTACGCAAGAGAACTAACCAAGCTGAAATCTCAGAGAACTTTGCTTTAC 2480
Db 2341 GTTCATGTACGCAAGAGAACTAACCAAGCTGAAATCTCAGAGAACTTTGCTTTAC 2400
Qy 2481 TAAGCTGAGTCACTTGAGAGCAATCTTAACCAATGCGGCACTGTAGTGGCTGTTTC 2540
Db 2401 TAAGCTGAGTCACTTGAGAGCAATCTTAACCAATGCGGCACTGTAGTGGCTGTTTC 2460
Qy 2541 TACCACATGACTTTTAAACATGTTTATATCAATTTTAAATTTTATGATACGCTAGTCTC 2600
Db 2461 TACCACATGACTTTTAAACATGTTTATATCAATTTTAAATTTTATGATACGCTAGTCTC 2520
Qy 2601 AGGAGAAATGTAATTTTCTATGTAATGAAATTCCTTTTCAAGTTTGTTCATTATAACAGT 2660
Db 2521 AGGAGAAATGTAATTTTCTATGTAATGAAATTCCTTTTCAAGTTTGTTCATTATAACAGT 2580
Qy 2661 TATTAATTTTAAATCAGCGTTAGAGTTTGTGCTGCTGCAACTGCTGTGAAAATTTCTCTGA 2720
Db 2581 TATTAATTTTAAATCAGCGTTAGAGTTTGTGCTGCTGCAACTGCTGTGAAAATTTCTCTGA 2640
Qy 2721 GTAAATCTGATTTTGTGAATGATCCAGACCAACCTCAGATTTTGTCAACCTGATTAAGT 2780
Db 2641 GTAAATCTGATTTTGTGAATGATCCAGACCAACCTCAGATTTTGTCAACCTGATTAAGT 2700
Qy 2781 CAATATGAATGATTTAAAGATGTGAGAACCAAAAAA 2825
Db 2701 CAATATGAATGATTTAAAGATGTGAGAACCAAAAAA 2745

RESULT 2
AK024865
LOCUS AK024865 2343 bp mRNA linear PRI 29-SEP-2000
DEFINITION Homo sapiens cDNA: FLJ21212 fis, clone COL00502.
ACCESSION AK024865
VERSION AK024865 1 GI:10437273
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (site)
Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
NEDO human cDNA sequencing project
Unpublished
```

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## FEATURES

## source

## CDS

## BASE COUNT

## ORIGIN

## Query Match

Best Local Similarity 82.1%; Score 2341; DB 9; Length 2343;

Matches 2338; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

```
Qy 508 CATTTTTATAATCAAGCTGGTCAACTCTCTTCGACAGTTTACAGTGTCTTTGAGACAT 567
Db 1 CATTTTTATAATCAAGCTGGTCAACTCTCTTCGACAGTTTACAGTGTCTTTGAGACAT 60
Qy 569 CCCCAGATATCTGCTAGAGAAGTGATCTTTGTAGATGACTACAGTGTATAGAGAGCACC 627
Db 61 CCCCAGATATCTGCTAGAGAAGTGATCTTTGTAGATGACTACAGTGTATAGAGAGCACC 120
Qy 628 TGAAGGAGCGCTTGGCCAAATGAGCTTTTCGGAGCTGCCCAAGGTGCGCTGTATCCCGCCA 687
Db 121 TGAAGGAGCGCTTGGCCAAATGAGCTTTTCGGAGCTGCCCAAGGTGCGCTGTATCCCGCCA 180
Qy 688 ACAAGAGAGAGCGCTGCTGCGAGCCCGCTGCTGGGGCGCTCTCGCGGAGGGCGGATG 747
Db 181 ACAAGAGAGAGCGCTGCTGCGAGCCCGCTGCTGGGGCGCTCTCGCGGAGGGCGGATG 240
Qy 748 TTCTGACCTTCTGAGCTGTCACTGTGAGTGCACCAAGAGGTGGCTGGAGCGCTGTCTGC 807
Db 241 TTCTGACCTTCTGAGCTGTCACTGTGAGTGCACCAAGAGGTGGCTGGAGCGCTGTCTGC 300
Qy 808 AGAGGATCCATGAAGAGAGTCCGAGTGGTGTGTCGCCGGTGAATGTATGATCGACTGGA 867
Db 301 AGAGGATCCATGAAGAGAGTCCGAGTGGTGTGTCGCCGGTGAATGTATGATCGACTGGA 360
Qy 868 ACACCTTCGAATACCTGGGAACCTCCGGGAGCCCGAGATCGGCGGTTTTCAGCTGGAGGC 927
Db 361 ACACCTTCGAATACCTGGGAACCTCCGGGAGCCCGAGATCGGCGGTTTTCAGCTGGAGGC 420
Qy 928 TGGTGTTCACCTGGCACACAGTTCTCTGAGAGGAGAGATACGGATGCAATCCCGCGTCG 987
```

2 (bases 1 to 2343)

Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,  
Shibahara,T., Tanaka,T. and Nakamura,Y.

Direct Submission

Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,  
University of Tokyo, Laboratory of Genome Structure Analysis, Human  
Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,  
Japan (E-mail:cdnal@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,  
Fax:81-3-5449-5416)NEDO human cDNA sequencing project supported by Ministry of  
International Trade and Industry of Japan; cDNA full insert  
sequencing; Research Association for Biotechnology; cDNA library  
construction, 5'- & 3'-end one pass sequencing; Department of  
Virology and Human Genome Center, Institute of Medical Science,  
University of Tokyo (partly supported by Science and Technology  
Agency).

## Location/Qualifiers

1..2343

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="COL00502"

/tissue\_type="colon"

/clone\_lib="COL"

/notes="cloning vector pME18SFL3"

465..1319

/notes="unlabeled protein product"

/codon\_start=1

/protein\_id="BAB15027.1"

/db\_xref="GI:10437274"

translation="MQSPDVIRPTMAGGLFAVSKYFEYLGSDYDTGNEVNGENLE  
FSFRWOCGVLETHPCSHVGHVFPKQAPYSRNKALANSVRAAEVWMDFFKELYHRN  
PBRLPEPDVTERKQLDKLOCKDFKNFLVTVBELHVPEDRPGFGLMKNLGLTDY  
CDYNPBENQIYLGHQVILYLCHGMQNGOFFEYTSKEIRYNTHQPEGCIAYEAGMDT  
LIMHLCBETAPENQKFILOEDSLFHEQSKKCVQAAKRESSDSFVPLLRDCTNSDQK  
WFFKERML"

673 a 474 c 563 g 633 t

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 7, 2003, 12:45:48 ; Search time 10396.1 Seconds  
(without alignments)  
11215.070 Million cell updates/sec

Title: US-10-074-527-1

Perfect score: 2850

Sequence: 1 ccggctcggtaccactataa.....aaaaaaaaaaaaaaaaaaaaa 2850

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:

1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vi:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rnd:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_htg\_hum:\*  
40: em\_htg\_mus:\*  
41: em\_htg\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2738.2	96.1	2745	6	AX477705 Sequence
2	2341	82.1	2343	9	AK024865 Homo sapi
3	1838.6	64.5	1928	9	BC013945 Homo sapi
4	1744.4	61.2	1746	9	AB078146 Homo sapi
5	1744.4	61.2	1746	9	AB078146 Homo sapi
6	1127	39.5	149930	9	HSAL132365
7	1053.2	37.0	1808	9	AL136084
8	1039	36.5	52737	2	AC007800 Homo sapi
9	921.6	32.3	63632	2	AC132018 4
10	921.6	32.3	110000	2	AC098559-2
11	911.4	32.0	288118	2	AC129139 Rattus no
12	907.6	31.8	228368	2	AC121727
13	906	31.8	288118	2	AC129139
14	887.4	31.1	110000	2	AC099237 3
15	887.4	31.1	166547	2	AC134060
16	873.4	30.6	110000	2	AC098559 1
17	870.4	30.6	166547	2	AC134060
18	865	30.4	110000	2	AC095991 0
19	865	30.4	110000	2	AC112324 0
20	865	30.4	110000	2	AC115635 2
21	865	30.4	299134	2	AC128783
22	853	29.9	110000	2	AC105836 1
23	844.8	29.6	94616	2	AC098559 7
24	550	19.3	110000	2	AC099237 0
25	527	18.5	554	6	AX197983 Sequence
26	526.4	18.5	544	6	AX208574 Sequence
27	526.2	18.5	328901	2	AC130864 Rattus no
28	522.4	18.3	226623	2	AC097343
29	520.8	18.3	1737	10	MMU73819
30	520.8	18.3	234109	2	AC101864
31	499.4	17.5	5369	9	BC036390 Homo sapi
32	499.4	17.5	73481	9	AC025034 Homo sapi
33	497	17.4	852	6	AX197725 Sequence
34	496.4	17.4	632	6	AX208332
35	496.4	17.4	632	6	AX209583 Sequence
36	496.2	17.4	1737	6	AR236647 Sequence
37	496.2	17.4	1737	9	HSY08564
38	493.6	17.3	165175	2	AC083808
39	409	14.4	421	6	AX330041
40	409	14.4	421	6	AX335657
41	363	12.7	483	6	AX371123
42	300.8	10.6	2384	3	AK112431
43	285.6	10.0	1920	6	AR153422
44	285.6	10.0	2575	6	AX376296
45	285.6	10.0	2575	6	AX697278 Sequence

# ALIGNMENTS

RESULT 1  
AX477705  
LOCUS AX477705 2745 bp DNA linear PAT 12-AUG-2002  
DEFINITION Sequence 17 from Patent WO0246426.  
ACCESSION AX477705  
VERSION AX477705.1 GI:22216860  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Sanjanwala,M.M., Yao,M.G., Au-Young,J., Baughn,M.R., Arvizu,C.,  
Ring,H.Z., Lee,E.A., Ding,L., Hafalia,A.J., Tang,Y.T., Yue,H.,  
Tribouley,C.M., Lu,D.A., Lai,P.G., Warren,B.A., Yang,J.,

Walia,N.K., Nguyen,D.B., Gandhi,A.R. and Ison,C.H.

Drug metabolizing enzymes  
Patent: WO 0246426-A 17 13-JUN-2002;  
Incyte Genomics, Inc. (US)

Location/Qualifiers

1. .2745

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/note="Incyte ID No: 2860635CB1"

704 a 612 c 737 g 692 t

BASE COUNT  
ORIGIN

Query Match 96.1%; Score 2738.2; DB 6; Length 2745;

Best Local Similarity 99.7%; Pred. No. 0;

Matches 2737; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Qy	81	ATGTGGGGGCGCAGCGCGCGCGCGCTGCCCGCGGAACTGCGGCGCGCGCGCGGAGCG	140
Db	1	ATGTGGGGGCGCAGCGCGCGCGCGCTGCCCGCGGAACTGCGGCGCGCGCGGAGCG	60
Qy	141	CTGTTGTTGCTCTGGCGCTACTGGCGTTGGCCGGGCTGGGCTGGGCTGGGCGGCAG	200
Db	61	CTGTTGTTGCTCTGGCGCTACTGGCGTTGGCCGGGCTGGGCTGGGCTGGGCGGCAG	120
Qy	201	CGTGGGGCGGGCGCGGGCTGCCGAGCCGGAGCCCGCGCGCACCCCGCGCCCGCGGCGG	260
Db	121	CGTGGGGCGGGCGGGGCTGCCGAGCCGGAGCCCGCGCACCCCGCGCCCGCGGCGG	180
Qy	261	CGCGAGCGGTCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG	320
Db	181	CGCGAGCGGTCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG	240
Qy	321	CGCGTGGCGGTCAGCTGCGAGGGCGAGAGCTGGCGCTGCGAGGAGAGCGTGGCGCTG	380
Db	241	CGCGTGGCGGTCAGCTGCGAGGGCGAGAGCTGGCGCTGCGAGGAGAGCGTGGCGCTG	300
Qy	381	CACAGATTAACTACTACCTCAGCGCGCGCTCACTGCGACCGCGCGCGCGCGCGCGC	440
Db	301	CACAGATTAACTACTACCTCAGCGCGCGCTCACTGCGACCGCGCGCGCGCGCGC	360
Qy	441	TGGAACCGGCTGTGCAAGAGAGAAATATGATTATGATTAATTTGCCCAGGACATCTGTT	500
Db	361	TGGAACCGGCTGTGCAAGAGAGAAATATGATTATGATTAATTTGCCCAGGACATCTGTT	420
Qy	501	ATCATAGCATTTTATATGAAGCCTGGTCAACTCTCTTGGACAGTTTACAGTGTCTTT	560
Db	421	ATCATAGCATTTTATATGAAGCCTGGTCAACTCTCTTGGACAGTTTACAGTGTCTTT	480
Qy	561	GAGACATCCCGGATATCTGCTAGAAGAGTGTATCTTGTAGATGACTACAGTGATAGA	620
Db	481	GAGACATCCCGGATATCTGCTAGAAGAGTGTATCTTGTAGATGACTACAGTGATAGA	540
Qy	621	GAGCACCCTGAAGGAGCGCTTGGCAATGAGCTTTGGGAGCTGCCCAAGGTCGCGCTGATC	680
Db	541	GAGCACCCTGAAGGAGCGCTTGGCAATGAGCTTTGGGAGCTGCCCAAGGTCGCGCTGATC	600
Qy	681	CGCGCCAAAGAGAGAGGCGCTGGTGGAGCCCGGCTGCTGGGGGCGTCTGGCGCGAGG	740
Db	601	CGCGCCAAAGAGAGAGGCGCTGGTGGAGCCCGGCTGCTGGGGGCGTCTGGCGCGAGG	660
Qy	741	GGCGATGTTCTGACCTTCTGGAATGCTCACTGTGAGTGCACGAAGGCTGGCGAGCGG	800
Db	661	GGCGATGTTCTGACCTTCTGGAATGCTCACTGTGAGTGCACGAAGGCTGGCGAGCGG	720
Qy	801	CTGCTGCAGAGGATCCATGAGAGAGGTCGCGAGTGGTGGTGGCGGCTGATTGATGTCATC	860
Db	721	CTGCTGCAGAGGATCCATGAGAGAGGTCGCGAGTGGTGGTGGCGGCTGATTGATGTCATC	780
Qy	861	GACTGGAAACCTTTGGAATACCTGGGAACTCCCGGGAGGCCAGATCGCGGTTTCGAC	920
Db	781	GACTGGAAACCTTTGGAATACCTGGGAACTCCCGGGAGGCCAGATCGCGGTTTCGAC	840

Qy	921	TGGAGGCTGGTGTTCACGTGGCACACAGTTCTTGAGAGGAGAGGATACGGATGCAATCC	980
Db	841	TGGAGGCTGGTGTTCACGTGGCACACAGTTCTTGAGAGGAGAGGATACGGATGCAATCC	900
Qy	981	CCCGTCGATGTCATCAGGTCCTCAACAATGGCTGGTGGGCTGTTTGTGTGAGTAAGAAA	1040
Db	901	CCCGTCGATGTCATCAGGTCCTCAACAATGGCTGGTGGGCTGTTTGTGTGAGTAAGAAA	960
Qy	1041	TATTTTGAATATCTGGGGCTTTATGATACAGGAATGGAAGTTTGGGGAGGAGAAAACCTC	1100
Db	961	TATTTTGAATATCTGGGGCTTTATGATACAGGAATGGAAGTTTGGGGAGGAGAAAACCTC	1020
Qy	1101	GAATTTTCTTTAGGATCTGGCAGTGTGGTGGGTTCTGGAACAACACACCCATGTTCCCAT	1160
Db	1021	GAATTTTCTTTAGGATCTGGCAGTGTGGTGGGTTCTGGAACAACACACCCATGTTCCCAT	1080
Qy	1161	GTTGGGCATGTTTTCCTCCCAAGCAAGCTCCCTACTCCCGCAACAAGGCTCTGGCCAAAGT	1220
Db	1081	GTTGGGCATGTTTTCCTCCCAAGCAAGCTCCCTACTCCCGCAACAAGGCTCTGGCCAAAGT	1140
Qy	1221	GTTTCGTGACGCTGAAGTATGGATGGAATTTAAAGAGCTCTACTACCATCGCAACCCC	1280
Db	1141	GTTTCGTGACGCTGAAGTATGGATGGAATTTAAAGAGCTCTACTACCATCGCAACCCC	1200
Qy	1281	CGTGCCTCGCTTGGAACTTTTGGGGATGTGACAGAGAGGAAGCAGCTCCGGGACAAGCTC	1340
Db	1201	CGTGCCTCGCTTGGAACTTTTGGGGATGTGACAGAGAGGAAGCAGCTCCGGGACAAGCTC	1260
Qy	1341	CAGTGTAAAGACTTCAAGTGTCTTGGAGACTGTGTATCCAGAACTGCAATGTCCTGAG	1400
Db	1261	CAGTGTAAAGACTTCAAGTGTCTTGGAGACTGTGTATCCAGAACTGCAATGTCCTGAG	1320
Qy	1401	GACAGCTGGCTTCTTCGGGATGCTCCAGAACAAAGGACTTAACAGACTACTGCTTTGAC	1460
Db	1321	GACAGCTGGCTTCTTCGGGATGCTCCAGAACAAAGGACTTAACAGACTACTGCTTTGAC	1380
Qy	1461	TATAACCTCCCGATGAAAACACAGATTGTGGGACACAGAGTCACTTGTACTCTGTCTAT	1520
Db	1381	TATAACCTCCCGATGAAAACACAGATTGTGGGACACAGAGTCACTTGTACTCTGTCTAT	1440
Qy	1521	GGATGGGCGCAGAACTCAGTTTTTTCAGTACACGTCCAGAAAGAAATACGCTATAACACC	1580
Db	1441	GGATGGGCGCAGAACTCAGTTTTTTCAGTACACGTCCAGAAAGAAATACGCTATAACACC	1500
Qy	1581	CACAGCCTGAGGCTGCAATGCTGTGGAGAGAGGATGATACCTTATCATGTCATCTC	1640
Db	1501	CACAGCCTGAGGCTGCAATGCTGTGGAGAGAGGATGATACCTTATCATGTCATCTC	1560
Qy	1641	TGCGAAGAAACTGCCCCAGAGAACTCAGAACTTTCATCTTGAGAGGATGGATCTTTATTT	1700
Db	1561	TGCGAAGAAACTGCCCCAGAGAACTCAGAACTTTCATCTTGAGAGGATGGATCTTTATTT	1620
Qy	1701	CAGAAACAGTCCAAAGAAATGTGTCCAGGCTGCGAGGAGAGTGCAGTGCAGTTTCGTT	1760
Db	1621	CAGAAACAGTCCAAAGAAATGTGTCCAGGCTGCGAGGAGAGTGCAGTGCAGTTTCGTT	1680
Qy	1761	CCACTCTTACGAGACTGCACCAACTCGGATCATCAGAAATGGTTCCTTCAAGAGCGCATG	1820
Db	1681	CCACTCTTACGAGACTGCACCAACTCGGATCATCAGAAATGGTTCCTTCAAGAGCGCATG	1740
Qy	1821	TTATGAAGCCTCGTGTATCAAGGAGCCCATCGAAGGAGACTGTGGAGCCAGGACTCTGCC	1880
Db	1741	TTATGAAGCCTCGTGTATCAAGGAGCCCATCGAAGGAGACTGTGGAGCCAGGACTCTGCC	1800
Qy	1881	CAACAAAGACTTACGTAAAGCAGTGACAGAAACCCACCAAACTAGGCTGCATTGCTTGG	1940
Db	1801	CAACAAAGACTTACGTAAAGCAGTGACAGAAACCCACCAAACTAGGCTGCATTGCTTGG	1860
Qy	1941	AAGAGCAATCATTTTGCCTATTTGGAAGTGTGTGGATTTAGTAAATATGTAATAA	2000
Db	1861	AAGAGCAATCATTTTGCCTATTTGGAAGTGTGTGGATTTAGTAAATATGTAATAA	1920
Qy	2001	GCTTTGTACTTATTTTGGAGAACTTTTTTAAATGTTTCCAAATAACCTTATTTTCAAGGGTA	2060

```
Db 1921 GCCTTGACTATTTTGGAGAACTTTTAAAGTTCCTAAATACCCCTATTTTCAAAGGGTA 1980
Qy 2061 ATCGTAAGATGTTAAACCTTTGGTATTTAGAAAATTTAAACCTTATATATTTTCTATCA 2120
Db 1981 ATCGTAAGATGTTAAACCTTTGGTATTTAGAAAATTTAAACCTTATATATTTTCTATCA 2040
Qy 2121 ABAWRWAWATTTTACAGTCTGCGCTTTTACTCTCTCAATTAGCAAAAAGATAAAGATTTTAT 2180
Db 2041 AGATGTATATTTTACAGTCTGCGCTTTTACTCTCTCAATTAGCAAAAAGATAAAGATTTTAT 2100
Qy 2181 TTTGGTATTTTACAAGAAATCCAGGTACGAAGATATCTGATGGTGGGATCAAGTTCAGGTTCA 2240
Db 2101 TTTGGTATTTTACAAGAAATCCAGGTACGAAGATATCTGATGGTGGGATCAAGTTCAGGTTCA 2160
Qy 2241 AGCAACGTACTTTGCATTAATCACTGATTAATCACTCAGCTGCGGGGTTAAAGTTTCCCAAGTA 2300
Db 2161 AGCAACGTACTTTGCATTAATCACTGATTAATCACTCAGCTGCGGGGTTAAAGTTTCCCAAGTA 2220
Qy 2301 TAGAGAGACTGTCACTAGGAACATTTGATTTATTTAGTTCATTTAGGTCATTTGAGATCTTTAGA 2360
Db 2221 TAGAGAGACTGTCACTAGGAACATTTGATTTATTTAGTTCATTTAGGTCATTTGAGATCTTTAGA 2280
Qy 2361 TGTATTTTAAAGAAAGTCTTTTGGTATGTTGCTTACCAAGTAACTTAACTCCATTAAT 2420
Db 2281 TGTATTTTAAAGAAAGTCTTTTGGTATGTTGCTTACCAAGTAACTTAACTCCATTAAT 2340
Qy 2421 GTTCATGTCAGCCAAAGAGACTAAACCAAGCTGAAATCTCAGAGAACAAATTTGCTTTTAC 2480
Db 2341 GTTCATGTCAGCCAAAGAGACTAAACCAAGCTGAAATCTCAGAGAACAAATTTGCTTTTAC 2400
Qy 2481 TAAGCTGAGTCAACTGAGAGAGAACTTTTAAACAACTGCCAGCTGTAGTGTGCTGTTTC 2540
Db 2401 TAAGCTGAGTCAACTGAGAGAGAACTTTTAAACAACTGCCAGCTGTAGTGTGCTGTTTC 2460
Qy 2541 TACCACATGACATTTAAACATGTTTATATCAATTTTAAATTTTATGATACGCTAGTGTTC 2600
Db 2461 TACCACATGACATTTAAACATGTTTATATCAATTTTAAATTTTATGATACGCTAGTGTTC 2520
Qy 2601 AGGGAGAAATGTAATGTTTATATGAAATTCCTTTTCAAAGTTGTTTCATTAATAACAGT 2660
Db 2521 AGGGAGAAATGTAATGTTTATATGAAATTCCTTTTCAAAGTTGTTTCATTAATAACAGT 2580
Qy 2661 TATTAATTTAAATCAGGTTAGAGTTTGTGCTGCTGCAACTGCTGTGAAAATTTCTCTGA 2720
Db 2581 TATTAATTTAAATCAGGTTAGAGTTTGTGCTGCTGCAACTGCTGTGAAAATTTCTCTGA 2640
Qy 2721 GTAAATCTGATTTGTGAATGATCCAGACCAACCTGAGATTTTGTCAACCTGATTAAGT 2780
Db 2641 GTAAATCTGATTTGTGAATGATCCAGACCAACCTGAGATTTTGTCAACCTGATTAAGT 2700
Qy 2781 CAATATGAATGATTTAAAGAGATGTGAGAACAAAAA 2825
Db 2701 CAATATGAATGATTTAAAGAGATGTGAGAACAAAAA 2745
```

```
RESULT 2
AK024865
LOCUS Homo sapiens cDNA: FLJ21212 fis, clone COL00502. 2343 bp mRNA linear PRI 29-SEP-2000
DEFINITION AK024865
ACCESSION AK024865
VERSION AK024865.1 GI:10437273
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (ntese)
Kawabata, A., Hikiji, T., Kobatake, N., Inagaki, H., Ikema, Y.,
Okamoto, S., Okitani, R., Ota, T., Suzuki, Y., Obayashi, M., Nishi, T.,
Shibahara, T., Tanaka, T., Nakamura, Y., Isogai, T. and Sugano, S.
NEDO human cDNA sequencing project
Unpublished
```

REFERENCE  
AUTHORSTITLE  
JOURNAL

## COMMENT

FEATURES  
source

## CDS

BASE COUNT  
ORIGIN

Query Match 82.1%; Score 2341; DB 9; Length 2343;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 2338; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

```
Qy 508 CATTTTATTAATGAAGCCTGGTCAACTCTCTCGACAGTTTACAGTGTCTTTGAGACAT 567
Db 1 CATTTTATTAATGAAGCCTGGTCAACTCTCTCGACAGTTTACAGTGTCTTTGAGACAT 60
Qy 568 CCCCAGATATCTGTAGAGAAGTGTATCTTGTAGATGACTACAGTATAGAGAGACC 627
Db 61 CCCCAGATATCTGTAGAGAAGTGTATCTTGTAGATGACTACAGTATAGAGAGACC 120
Qy 628 TGAAGGAGCGCTTGGCCAAATGAGCTTTCCGGAGTCCCAAGGTGCGCTGATCCGCCCA 687
Db 121 TGAAGGAGCGCTTGGCCAAATGAGCTTTCCGGAGTCCCAAGGTGCGCTGATCCGCCCA 180
Qy 688 ACAAGAGAGAGCGCTTGGTCCGAGCCCGCTGTGCGGGCGCTCTGCGGCGAGGGCGCATG 747
Db 181 ACAAGAGAGAGCGCTTGGTCCGAGCCCGCTGTGCGGGCGCTCTGCGGCGAGGGCGCATG 240
Qy 748 TTCTGACCTTCTGGACTGTCTAGTGTGAGTGCACCAAGGGTGGCTGGAGCCGCTGCTGC 807
Db 241 TTCTGACCTTCTGGACTGTCTAGTGTGAGTGCACCAAGGGTGGCTGGAGCCGCTGCTGC 300
Qy 808 AGAGGATCCATGAAGAGAGTGGCAGTGGTGGCCGGTGTGATGTGATCGATCGATCGGA 867
Db 301 AGAGGATCCATGAAGAGAGTGGCAGTGGTGGCCGGTGTGATGTGATCGATCGATCGGA 360
Qy 868 ACACCTTCCAAATACCTGGGGAACCTCCGGGAGGCCCCAGATCGCGGCTTTTCAGCTGGAGGC 927
Db 361 ACACCTTCCAAATACCTGGGGAACCTCCGGGAGGCCCCAGATCGCGGCTTTTCAGCTGGAGGC 420
Qy 928 TGGTGTTCACGTGGGCAACAGTTCTTGAGAGGAGAGGATACGGATGCAATCCCCGCTCG 987
```

2 (bases 1 to 2343)  
Sugano, S., Suzuki, Y., Ota, T., Obayashi, M., Nishi, T., Isogai, T.,  
Shibahara, T., Tanaka, T. and Nakamura, Y.  
Direct Submission  
Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,  
University of Tokyo, Laboratory of Genome Structure Analysis, Human  
Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,  
Japan (E-mail:cdnal@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,  
Fax:81-3-5449-5416)  
NEDO human cDNA sequencing project supported by Ministry of  
International Trade and Industry of Japan; cDNA full insert  
sequencing; Research Association for Biotechnology; cDNA library  
construction, 5'- & 3'-end one pass sequencing; Department of  
Virology and Human Genome Center, Institute of Medical Science,  
University of Tokyo (partly supported by Science and Technology  
Agency).

## Location/Qualifiers

1. 2343  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="COL00502"  
/tissue\_type="colon"  
/clone\_lib="COL"  
/notes="cloning vector pME18SFL3"  
465..1319  
/notes="unnamed protein product"  
/codon\_start=1  
/protein\_id="BAB15027.1"  
/db\_xref="GI:10437274"

BASE COUNT 673 a 474 c 563 g 633 t

Db 421 TGGTGTTCACGTGGCACACACAGTTCCTGAGAGGAGAGGATACGGATGCAATCCCGCTCG 480  
Qy 988 ATGTCATCAGGTCTCCAAACAATGGCTGGCTGGCTGTTGCTGTGAGTAAAGAAATATTTG 1047  
Db 481 ATGTCATCAGGTCTCCAAACAATGGCTGGCTGGCTGTTGCTGTGAGTAAAGAAATATTTG 540  
Qy 1048 AATATCTGGGGTCTTATGATACAGGAATGAAGTTTGGGGAGGAGAAAACCTCGAATTTT 1107  
Db 541 AATATCTGGGGTCTTATGATACAGGAATGAAGTTTGGGGAGGAGAAAACCTCGAATTTT 600  
Qy 1108 CTTTTAGGATCTGGCAGTGTGGGTTCGAAACACACACCATGTTCCATGTTGGCC 1167  
Db 601 CTTTTAGGATCTGGCAGTGTGGGTTCGAAACACACACCATGTTCCATGTTGGCC 660  
Qy 1168 ATGTTTTCCCAAGCAGCTCCCTACTCCCGCAACAGGCTCTGGCCACAGTGTTCGTG 1227  
Db 661 ATGTTTTCCCAAGCAGCTCCCTACTCCCGCAACAGGCTCTGGCCACAGTGTTCGTG 720  
Qy 1228 CAGCTGAAGTATGATGGATGAATTTAAAGAGCTCTACTACCATCGCAACCCCGTGCCC 1287  
Db 721 CAGCTGAAGTATGATGGATGAATTTAAAGAGCTCTACTACCATCGCAACCCCGTGCCC 780  
Qy 1288 GCTTGGAAACCTTTTGGGGATGTGACAGAGAGAGCAGCTCCGGGACAAAGCTCCAGTGA 1347  
Db 781 GCTTGGAAACCTTTTGGGGATGTGACAGAGAGAGCAGCTCCGGGACAAAGCTCCAGTGA 840  
Qy 1348 AAGACTTCAAGTGGTCTTGGAGACTGTGTATCCAGAACTGCATGTGCTTGAGGACAGGC 1407  
Db 841 AAGACTTCAAGTGGTCTTGGAGACTGTGTATCCAGAACTGCATGTGCTTGAGGACAGGC 900  
Qy 1408 CTGGCTTCTTCGGGATGCTCCAGAAACAAAGAGCTTAACAGACTACTGCTTTGACTATAACC 1467  
Db 901 CTGGCTTCTTCGGGATGCTCCAGAAACAAAGAGCTTAACAGACTACTGCTTTGACTATAACC 960  
Qy 1468 CTCCCGATGAAAACAGATGTGGGACACACAGGTCACTTGTACCTCTGTCTGGGATGG 1527  
Db 961 CTCCCGATGAAAACAGATGTGGGACACACAGGTCACTTGTACCTCTGTCTGGGATGG 1020  
Qy 1528 GCCAGATCAGTTTTTCGAGTACAGTCCAGAAAGAAATACGCTATTAACACCCACCAGC 1587  
Db 1021 GCCAGATCAGTTTTTCGAGTACAGTCCAGAAAGAAATACGCTATTAACACCCACCAGC 1080  
Qy 1588 CTGAGGCTGCTGATGCTGTGAAGCAGAAATGGAATACCTTTATCATGCTACTCTGCGAAG 1647  
Db 1081 CTGAGGCTGCTGATGCTGTGAAGCAGAAATGGAATACCTTTATCATGCTACTCTGCGAAG 1140  
Qy 1648 AAATGCCCCAGAGAATCAGAAATTCATCTTTCAGGAGGATGGATCTTTATTTTACGAAAC 1707  
Db 1141 AAATGCCCCAGAGAATCAGAAATTCATCTTTCAGGAGGATGGATCTTTATTTTACGAAAC 1200  
Qy 1708 AGTCCAGAAATGTGTCAGGCTCGAGGAGGAGGAGTGCAGTGACAGTTCGTTCCACTCT 1767  
Db 1201 AGTCCAGAAATGTGTCAGGCTCGAGGAGGAGGAGTGCAGTGACAGTTCGTTCCACTCT 1260  
Qy 1768 TACGAGACTGCACCAACTCGGATCATCAGAAATGTTCTTCAAAGAGCGATGTTATGAA 1827  
Db 1261 TACGAGACTGCACCAACTCGGATCATCAGAAATGTTCTTCAAAGAGCGATGTTATGAA 1320  
Qy 1828 GCCTCGTGTATCAAGAGGCCCATCGAAGGAGACTGTGGAGCGAGGACTCTGCCCAACAAA 1887  
Db 1321 GCCTCGTGTATCAAGAGGCCCATCGAAGGAGACTGTGGAGCGAGGACTCTGCCCAACAAA 1380  
Qy 1888 GACTTAGTAAAGAGTGAACAGAACCCCAACAAACCTAGGCTGATGCTTTGAGAGGC 1947  
Db 1381 GACTTAGTAAAGAGTGAACAGAACCCCAACAAACCTAGGCTGATGCTTTGAGAGGC 1440  
Qy 1948 AATCATTTGCGCATTTGTGAAGTGTGTTGGATTTAGTAAATGTAATGTAATGTAATGTA 2007  
Db 1441 AATCATTTGCGCATTTGTGAAGTGTGTTGGATTTAGTAAATGTAATGTAATGTAATGTA 1500  
Qy 2008 ACTTATTTGAGAACTTTTTTAAATGTTCCAAAATACCCCTATTTTCAAAGGGTAAATCGTAA 2067  
Db 1501 ACTTATTTGAGAACTTTTTTAAATGTTCCAAAATACCCCTATTTTCAAAGGGTAAATCGTAA 1560

Qy 2068 GATGTTAAACCCCTTGGTATTTAGAAAATTAACCTTTATAATATTTTCTATCAARAWRA 2127  
Db 1561 GATGTTAAACCCCTTGGTATTTAGAAAATTAACCTTTATAATATTTTCTATCAAGATGA 1620  
Qy 2128 WATTTTACAGTCTGCTGCTTTTACTCTCATAGCAAAAAGATAAAGATTTTATTTTGGTA 2187  
Db 1621 TATTTTACAGTCTGCTGCTTTTACTCTCATAGCAAAAAGATAAAGATTTTATTTTGGTA 1680  
Qy 2188 TTTTACAGAAATTTCCAGGTACGAAGATATCTGCATGGGTGGAATCAGGTTCAAGCAACG 2247  
Db 1681 TTTTACAGAAATTTCCAGGTACGAAGATATCTGCATGGGTGGAATCAGGTTCAAGCAACG 1740  
Qy 2248 TACTTTGCATTAACTGATAATACCTCAGCTGGGGGTAAAGTTTCCCCAGTATAGAGAG 2307  
Db 1741 TACTTTGCATTAACTGATAATACCTCAGCTGGGGGTAAAGTTTCCCCAGTATAGAGAG 1800  
Qy 2308 ACTGTCATAGGAACAATGATTTATTTAGTATTTTCAAGTCAATGAGATCTTCTAGATGATTT 2367  
Db 1801 ACTGTCATAGGAACAATGATTTATTTATTTTCAAGTCAATGAGATCTTCTAGATGATTT 1860  
Qy 2368 TAAAAAGAAATGCTTTTGGTATGTTGTTGCTACCAAGTAAACCTCCATAATGTTTCATG 2427  
Db 1861 TAAAAAGAAATGCTTTTGGTATGTTGTTGCTACCAAGTAAACCTCCATAATGTTTCATG 1920  
Qy 2428 TCAGCCAAAGAGGAGTAAACCAAGCTGAAATCTCAGAGAAACAATTTGCTTTACTTAAAGCTG 2487  
Db 1921 TCAGCCAAAGAGGAGTAAACCAAGCTGAAATCTCAGAGAAACAATTTGCTTTACTTAAAGCTG 1980  
Qy 2488 AGTCAAATTTGAGAGCGAACTTCTTAAACAATGCGCACTGTAGTGTGCTGGTCTTACCACT 2547  
Db 1981 AGTCAAATTTGAGAGCGAACTTCTTAAACAATGCGCACTGTAGTGTGCTGGTCTTACCACT 2040  
Qy 2548 ATGACTTTTAAACATGTTTATATATATTTTAAATTTTATGATACGCTAGTGTGAGGAGA 2607  
Db 2041 ATGACTTTTAAACATGTTTATATATATTTTAAATTTTATGATACGCTAGTGTGAGGAGA 2100  
Qy 2608 AATGTAATGTTCTTATATGAAATTCCTTTTCAAGTGTGTTTCAATTAATAACAGTATTAAT 2667  
Db 2101 AATGTAATGTTCTTATATGAAATTCCTTTTCAAGTGTGTTTCAATTAATAACAGTATTAAT 2160  
Qy 2668 TTAATCAGCGTTAGAGTTTGTGCTGCTGCAACTGCTGTGAAAATTTCTCTGAGTAATTC 2727  
Db 2161 TTAATCAGCGTTAGAGTTTGTGCTGCTGCAACTGCTGTGAAAATTTCTCTGAGTAATTC 2220  
Qy 2728 TGATTTGTGATGATCCAGACCAACCTGAGATTTTGTCAACCTGATTAAGTCAATATG 2787  
Db 2221 TGATTTGTGATGATCCAGACCAACCTGAGATTTTGTCAACCTGATTAAGTCAATATG 2280  
Qy 2788 AATGATTAAAAAGATGTGAGAACAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 2847  
Db 2281 AATGATTAAAAAGATGTGAGAACAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 2340  
Qy 2848 AAA 2850  
Db 2341 AAA 2343

RESULT 3  
BC013945  
LOCUS  
DEFINITION  
Homo sapiens, similar to hypothetical protein FLJ21212, clone  
MGC:24384 IMAGE:4064736, mRNA, complete cds.  
ACCESSION  
BC013945  
VERSION  
BC013945.1  
KEYWORDS  
MGC.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 1928)  
AUTHORS  
Straussberg, R.  
TITLE  
Direct Submission

## JOURNAL

Submitted (07-SEP-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK  
COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: [cgaps@emil.nih.gov](mailto:cgaps@emil.nih.gov)

Tissue Procurement: ATCC

cDNA Library Preparation: CLONTECH Laboratories, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Institute for Systems Biology

<http://www.systemsbio.org>

contact: [amadane@systemsbiology.org](mailto:amadane@systemsbiology.org)

Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia Greene, Mark Ketteman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAL Plate: 31 Row: p Column: 2  
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.

FEATURES  
source

1..1928

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="MGC:24384 IMAGE:4064736"

/tissue\_type="Bone marrow, chronic myelogenous leukemia"

/clone\_lib="NIH MGC\_54"

/lab\_host="DH10B"

/notes="Vector: pDNR-LIB"

93..911

/codon\_start=1

/product="Similar to hypothetical protein FLJ21212"

/protein\_id="AAH13945.1"

/db\_xref="GI:15530299"

/translation="MAGGLFAVSKYFEYLGSDYDGMVWGENLFFSPRIWCGGVLETHPCSHVGHVFQKQPYSRNKALANSVRAAYVMDFFKELYHRNPRARLEPFGDVT EKHQKDRDKQFPLETVYPPELVDPDRPGFGLQNKGLTDYCFDYNPPDENQI VGHQVLYLCHGQNGQFFRYTSQKEIRYNTHOPEGICIAVEAGMDTLIHLCEETAPE NQKFLIQEDSLFHEQSKCVQARKESSDSFVPLLRDCTNSDHQKFFKERM"

BASE COUNT 578 a 372 c 433 g 545 t

## ORIGIN

Query Match

Best Local Similarity 99.28; Pred. No. 1.6e-250; Length 1928;

Matches 1841; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY	989	TGTCATCAGCTCTCCAAATGGCTGGTGGGCTGTTTGTCTGTGAGTAAGAAATATTGA	1048
DB	74	TGCCATGTGGTCTCCAAATGGCTGGTGGGCTGTTTGTCTGTGAGTAAGAAATATTGA	133
QY	1049	ATATCTGGGCTCTATGATACAGGAATGGAAGTTGGGGAGGAGAAACCTCGAATTTTC	1108
DB	134	ATATCTGGGCTCTATGATACAGGAATGGAAGTTGGGGAGGAGAAACCTCGAATTTTC	193
QY	1109	CTTTAGGATCTGGCAGTGTGGTGGGTTCTGGAACACACCCATGTTCCCATGTTGGCCA	1168
DB	194	CTTTAGGATCTGGCAGTGTGGTGGGTTCTGGAACACACCCATGTTCCCATGTTGGCCA	253
QY	1169	TGTTTTCCCAAGCAAGCTCCCTACTCCCGCAACAAGGCTCTGGCCAAAGTGTCTGTGC	1228
DB	254	TGTTTTCCCAAGCAAGCTCCCTACTCCCGCAACAAGGCTCTGGCCAAAGTGTCTGTGC	313
QY	1229	AGCTGAAGTATGATGATGAATTAAGAGCTCTACTACATCGCAACCCCGTGGCCG	1288
DB	314	AGCTGAAGTATGATGATGAATTAAGAGCTCTACTACATCGCAACCCCGTGGCCG	373
QY	1289	CTTGGAACTTTTGGGATGTGACAGAGGAGGAGCAGCTCCGGGACAAAGCTCCAGTGTAA	1348
DB	374	CTTGGAACTTTTGGGATGTGACAGAGGAGGAGCAGCTCCGGGACAAAGCTCCAGTGTAA	433

Db	1514	CAGCCAAAGAGGACCTAAACCAAGCTGAATCTCTCAGAGAACAAATTTGCTTTTACTAAGCTGA	1573
Qy	2489	GTCAACTTTGAGAGCGAACTTCTAACAAATCGCGCACTGTAGTGTGGCTGGTTCTACCACTA	2548
Db	1574	GTCAACTTTGAGAGCGAACTTCTAACAAATCGCGCACTGTAGTGTGGCTGGTTCTACCACTA	1633
Qy	2549	TGACTTTAAACCATGTTTATATCATTTTTTAAATTTTATGATACGCTAGTGTCAAGGAGAA	2608
Db	1634	TGACTTTAAACCATGTTTATATCATTTTTTAAATTTTATGATACGCTAGTGTCAAGGAGAA	1693
Qy	2609	ATGTAATGTTCTATATGAAATTCCTTTTCAAGTTTGTTCATTAATTAACAGTTATTAAATT	2668
Db	1694	ATGTAATGTTCTATATGAAATTCCTTTTCAAGTTTGTTCATTAATTAACAGTTATTAAATT	1753
Qy	2669	TAAATCAGCGTTAGAGTTTGTGCTGTCGCAACTCTCTGGAATTTCTCTGAGTAATTTCT	2728
Db	1754	TAAATCAGCGTTAGAGTTTGTGCTGTCGCAACTCTCTGGAATTTCTCTGAGTAATTTCT	1813
Qy	2729	GATTTGTGAATGATGCCAGACCAACCTCTGAGATTTTGTCAACCTGATTAAGTCAATATGA	2788
Db	1814	GATTTGTGAATGATGCCAGACCAACCTCTGAGATTTTGTCAACCTGATTAAGTCAATATGA	1873
Qy	2789	ATGATTTAAAGATGTGAGAACAAAAAATAAAAAAAAAAAAAAAAAAAAAA	2843
Db	1874	ATGATTTAAAGATGTGAGAACACTCAAAAAAAAAAAAAAAAAAAAAA	1928
RESULT 4			
AB078146			
LOCUS	AB078146.1	GI:22122073	
DEFINITION	Homo sapiens GALNT12 mRNA for	1746 bp	linear
ACCESSION	UPP-N-acetyl-alpha-D-galactosamine:polypeptide		
VERSION	N-acetylgalactosaminyltransferase 12, complete cds.		
KEYWORDS	AB078146		
SOURCE	AB078146.1		
ORGANISM	Homo sapiens (human)		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	Guo, J.M., Zhang, Y., Chen, L., Iwasaki, H., Wang, H., Kubota, T., Tachibana, K. and Narimatsu, H.		
JOURNAL	Molecular cloning and characterization of a novel member of the		
MEDLINE	UPP-GalNAc:polypeptide N-acetylgalactosaminyltransferase family,		
PUBMED	PP-GalNAc-T12		
REFERENCE	FEBS Lett. 524 (1-3), 211-218 (2002)		
AUTHORS	22131447		
TITLE	12135769		
JOURNAL	2 (bases 1 to 1746)		
MEDLINE	Zhang, Y., Guo, J., Wan, H., Chen, L., Gotoh, M., Inaba, N., Kikuchi, N.,		
PUBMED	Direct Submission		
AUTHORS	Submitted (16-JAN-2002) Yan Zhang, Institute of Advanced Industrial		
TITLE	Science and Technology (AIST), Institute of Molecular and Cell		
JOURNAL	Biology (IMCB), Laboratory of Gene Function Analysis; Central-2,		
MEDLINE	1-1-1 Umazono, Tsukuba, Ibaraki 305-8568, Japan		
PUBMED	(E-mail: yan-zhang@aist.go.jp, Tel: 81-298-61-3197,		
AUTHORS	Fax: 81-298-61-3191)		
FEATURES	Location/Qualifiers		
source	1..1746		
gene	/organism="Homo sapiens"		
CDS	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/chromosome="9"		
	/map="9q22.33-22.31.3"		
	/tissue_type="lung"		
	1..1746		
	/gene="GALNT12"		
	1..1746		
	/gene="GALNT12"		
	/note="GalNAc-T12"		







repeat_region	16112..16355	/notes="LMC4 repeat: matches 7405..7639 of consensus"
repeat_region	16357..16767	/note="MT1F repeat: matches 116..512 of consensus"
repeat_region	16776..16869	/notes="MT1G repeat: matches 1..94 of consensus"
repeat_region	17401..17569	/notes="MIR repeat: matches 23..203 of consensus"
repeat_region	17690..18401	/note="LIPAL1 repeat: matches 5439..6153 of consensus"
misc_feature	17946..18202	/notes="Sequence from clone PCR only."
repeat_region	18487..18644	/notes="LMC5 repeat: matches 7656..7818 of consensus"
repeat_region	18645..18896	/note="AluX repeat: matches 1..251 of consensus"
misc_feature	18866..19033	/notes="Single clone region. Sequence from reads from a short insert library derived from a single pUC clone. Restriction digest data confirm the assembly."
repeat_region	18927..19232	/note="AluY repeat: matches 1..302 of consensus"
repeat_region	19269..19562	/notes="L2 repeat: matches 2389..2668 of consensus"
repeat_region	20429..20543	/notes="L2 repeat: matches 2315..2437 of consensus"
repeat_region	21245..21351	/notes="L2 repeat: matches 2192..2298 of consensus"
repeat_region	21956..22131	/notes="MT1J repeat: matches 74..264 of consensus"
repeat_region	22247..22449	/notes="MTR repeat: matches 28..243 of consensus"
repeat_region	22503..22743	/note="MTR repeat: matches 22..262 of consensus"
repeat_region	23280..23574	/notes="AluJo repeat: matches 1..295 of consensus"
repeat_region	23662..23878	/notes="MTR repeat: matches 29..262 of consensus"
repeat_region	24029..24092	/note="LMB8 repeat: matches 6109..6173 of consensus"
repeat_region	24631..24731	/notes="MTR repeat: matches 119..223 of consensus"
repeat_region	24774..25022	/notes="AluSp repeat: matches 66..305 of consensus"
repeat_region	25192..25436	/note="ORSL repeat: matches 244..467 of consensus"
repeat_region	26198..26248	/notes="MR5B repeat: matches 124..174 of consensus"
repeat_region	26261..26417	/notes="LMA2 repeat: matches 6151..6308 of consensus"
repeat_region	26418..26715	/note="AluX repeat: matches 1..299 of consensus"
repeat_region	26793..27075	/notes="AluX repeat: matches 1..283 of consensus"
repeat_region	27076..27694	/note="LMA2 repeat: matches 5522..6146 of consensus"
repeat_region	27734..27809	/note="MR5B repeat: matches 21..96 of consensus"
repeat_region	27875..27979	/notes="MR20 repeat: matches 26..131 of consensus"
repeat_region	28831..28969	/note="MTR repeat: matches 31..185 of consensus"
repeat_region	29761..29824	/note="LMB4 repeat: matches 6117..6175 of consensus"
repeat_region	29825..30131	/notes="AluJo repeat: matches 1..309 of consensus"
repeat_region	30132..30596	/note="LMB4 repeat: matches 5607..6117 of consensus"
repeat_region	30946..31288	/notes="L2 repeat: matches 1886..2323 of consensus"
repeat_region	31470..31509	/notes="MTR repeat: matches 105..144 of consensus"
repeat_region	31713..32007	

```
/note="AluSx repeat: matches 1. .293 of consensus"
repeat_region 32390. .32709
/note="L2 repeat: matches 2325. .2675 of consensus"
repeat_region 33291. .33322

Query Match
Best Local Similarity 97.4%; Pred. No. 5.8e-174;
Matches 1139; Conservative 5; Mismatches 25; Indels 0; Gaps 0;

Qy 1582 GGAGATGGATCTTTATTTACGAAACAGTCAAGAAATGTGTCAGGCTCGGAGGAGGA 1741
Db |||||
Qy 46535 GTAGATGGATCTTTATTTACGAAACAGTCAAGAAATGTGTCAGGCTCGGAGGAGGA 46594
Db |||||
Qy 1742 GTCAGATGACAGTTTCGTTCCACTCTTACGAGAGCTGCACCAACTCGGATCATCAGAAATG 1801
Db |||||
Qy 46595 GTCAGATGACAGTTTCGTTCCACTCTTACGAGAGCTGCACCAACTCGGATCATCAGAAATG 46654
Db |||||
Qy 1802 GTTCTTCAAGAGCGCATGTTATGAAGCCTCGTGTATCAAGGAGCCCATCGAAGGAGACT 1861
Db |||||
Qy 46655 GTTCTTCAAGAGCGCATGTTATGAAGCCTCGTGTATCAAGGAGCCCATCGAAGGAGACT 46714
Db |||||
Qy 1862 GTGAGCCAGGACTCTGCCCAAGAACTTTAGCTAAGCAGTGCACGAGACCCACCAAAA 1921
Db |||||
Qy 46715 GTGAGCCAGGACTCTGCCCAAGAACTTTAGCTAAGCAGTGCACGAGACCCACCAAAA 46774
Db |||||
Qy 1922 ACTAGGCTGCATTGCTTTGAAGAGGCAATCATTTTGGCATTGTGAAAGTTGTCTGGAT 1981
Db |||||
Qy 46775 ACTAGGCTGCATTGCTTTGAAGAGGCAATCATTTTGGCATTGTGAAAGTTGTCTGGAT 46834
Db |||||
Qy 1982 TTAGTAAAAATGTGAATAAGCTTTGTACTTTATTTTGAAGAACTTTTAAATGTTCCTAAAT 2041
Db |||||
Qy 46835 TTAGTAAAAATGTGAATAAGCTTTGTACTTTATTTTGAAGAACTTTTAAATGTTCCTAAAT 46894
Db |||||
Qy 2042 ACCCTATTTTCAAGGGTAATCGTAAGATGTTAACCTTGTGTTTGAAGAAATTTAAACC 2101
Db |||||
Qy 46895 ACCCTATTTTCAAGGGTAATCGTAAGATGTTAACCTTGTGTTTGAAGAAATTTAAACC 46954
Db |||||
Qy 2102 TTATAATATTTTCTATCAARAWAWATTTTACAGTCGTGCTCTTACTCTCATTTAGCA 2161
Db |||||
Qy 46955 TTATAATATTTTCTATCAAGATGTATTTTACAGTCGTGCTCTTACTCTCATTTAGCA 47014
Db |||||
Qy 2162 AAAAAGATAAAGATTTTATTTTGGTATTTTCAAGAAATTTCCAGGTACGAGATATCTGCA 2221
Db |||||
Qy 47015 AAAAAGATAAAGATTTTATTTTGGTATTTTCAAGAAATTTCCAGGTACGAGATATCTGCA 47074
Db |||||
Qy 2222 TGGGTGGAATCAGGTTCAAGCAACGACTTTTGCAATTAACGTGATAACCTCAGCTCGG 2281
Db |||||
Qy 47075 TGGGTGGAATCAGGTTCAAGCAACGACTTTTGCAATTAACGTGATAACCTCAGCTCGG 47134
Db |||||
Qy 2282 GGTAAAGTTTCCAGTATAGAGAGCTGTCAGTACGAGCAATTTGATTTTATTTTACG 2341
Db |||||
Qy 47135 GGTAAAGTTTCCAGTATAGAGAGCTGTCAGTACGAGCAATTTGATTTTATTTTACG 47194
Db |||||
Qy 2342 GTCATTGAGATCTTTCTAGATGATTTTAAAAAGAAATGCTTTTGGTATGTTGCTAC 2401
Db |||||
Qy 47195 GTCATTGAGATCTTTCTAGATGATTTTAAAAAGAAATGCTTTTGGTATGTTGCTAC 47254
Db |||||
Qy 2402 ACAGTTAACTCCATATATGTTTATGTCAGCAAGAGGACTTAACCAAGCTGAATCTC 2461
Db |||||
Qy 47255 ACAGTTAACTCCATATATGTTTATGTCAGCAAGAGGACTTAACCAAGCTGAATCTC 47314
Db |||||
Qy 2462 AGAGAACAAATTTGTTTACTAAGCTGAGTCAACTTTGAGAGGAACTTCTCAACATGCGC 2521
Db |||||
Qy 47315 AGAGAACAAATTTGTTTACTAAGCTGAGTCAACTTTGAGAGGAACTTCTCAACATGCGC 47374
Db |||||
Qy 2522 ACTGTAGTGGCTGGTTCTACCACTATGACTTTTAAAAATGCTTTTATATCATTTTTAAT 2581
Db |||||
Qy 47375 ACTGTAGTGGCTGGTTCTACCACTATGACTTTTAAAAATGCTTTTATATCATTTTTAAT 47434
Db |||||
Qy 2582 TTTATGATACGGTAGTGTACGGGAAATGTAATGTTCTATATGAATTCCTTTTCAAG 2641
Db |||||
Qy 47435 TTTATGATACGGTAGTGTACGGGAAATGTAATGTTCTATATGAATTCCTTTTCAAG 47494
Db |||||
```

```
Qy 2642 TTGTGTCATTAAATACAGTATTAAATTAATCAGCGTTAGAGTTGTGTGCTGCAACT 2701
Db |||||
Qy 47495 TTGTGTCATTAAATACAGTATTAAATTAATCAGCGTTAGAGTTGTGTGCTGCAACT 47554
Db |||||
Qy 2702 GCTGTGAAAAATTTCTCTGAGTAATTTCTGATTTGTGTAATGATCCAGACCAACCTTGAT 2761
Db |||||
Qy 47555 GCTGTGAAAAATTTCTCTGAGTAATTTCTGATTTGTGTAATGATCCAGACCAACCTTGAT 47614
Db |||||
Qy 2762 TTGTGCAACCTGATTAAGTCAATATGAATGATTAATAAAGATGTGAGAACAAAAA 2821
Db |||||
Qy 47615 TTGTGCAACCTGATTAAGTCAATATGAATGATTAATAAAGATGTGAGAACACTGACTGAG 47674
Db |||||
Qy 2822 AAAAAAATTTTCTGACTGTAACCAAAAAATAA 2850
Db |||||
Qy 47675 ATCTTTTGTACTGTAACCAAAAAATAA 47703
Db |||||

RESULT 7
LOCUS AB048901 1808 bp mRNA linear PRI 20-SEP-2000
DEFINITION Macaca fascicularis brain cDNA, clone:QnpA-17439.
ACCESSION AB048901
VERSION AB048901.1 GI:10241975
KEYWORDS fis (full insert sequence).
SOURCE Macaca fascicularis (crab-eating macaque)
ORGANISM Macaca fascicularis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
Cercopithecinae; Macaca.
REFERENCE 1 (sites)
AUTHORS Osada,N., Hida,M., Kusuda,J., Tanuma,R., Iseki,K., Hirai,M.,
Terao,K., Suzuki,Y., Sugano,S. and Hashimoto,K.
TITLE Isolation of full-length cDNA clones from macaque brain cDNA
libraries
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1808)
AUTHORS Hashimoto,K., Osada,N., Hida,M., Kusuda,J. and Sugano,S.
TITLE Direct Submission
JOURNAL Submitted (14-SEP-2000) Katsuyuki Hashimoto, National Institute of
Infectious Diseases, Division of Genetic Resources; 23-1, Toyama
1-chome, Shinjuku-ku, Tokyo 162-8640, Japan
[5-mail:khashim@nih.go.jp, URL:http://www.nih.go.jp/yoken/genebank/,
Tel:81-3-5285-1111(ex.2120), Fax:81-3-5285-1181]
Lab host: TOP10
Vector: pME18S-FL3 (Acc.No. AB009864)
R. Site1: DraIII (CACTGTGTG)
R. Site2: DraIII (CACTGTGTG)
Description: 1st strand cDNA was primed with an oligo (dT) primer
using specific 5' and 3' primers and amplified by PCR. The PCR
product was digested with SfiI and size selection was performed to
exclude fragments <1.5kb.The SfiI-digested PCR product was cloned
into distinct draIII sites of pME18S-FL3. XhoI sites just outside
the DraIII sites can be used to isolate the cDNA insert. Libraries
were constructed by Sugano et al.(University of Tokyo, Institute of
Medical Science). Custom primer used for sequencing
( 5' end primer [CTTCTGCTCTAAAGCTGCG];
3' end primer [CGACCTCAGCTCGACCA] ).
FEATURES
source
1. .1808
/organism="Macaca fascicularis"
/mol_type="mRNA"
/db_xref="taxon:9541"
/clone="QnpA-17439"
/sex="male"
/tissue_type="brain parietal lobe"
/clone_lib="macaque brain cDNA library QnpA"
/dev_stage="adult"
BASE COUNT 531 a 365 c 383 g 529 t
ORIGIN
Query Match 37.0%; Score 1053.2; DB 9; Length 1808;
Best Local Similarity 84.9%; Pred. No. 3.7e-162;
```

Matches 1300; Conservative 5; Mismatches 68; Indels 159; Gaps 5;

Qy 1292 GGAACTTTTGGGATGTGACAGAGAGAAAGCAGCTCCGGACAGAGCTCCAGTGTAAAGA 1351  
Db |||||  
Qy 421 GGAACCTTTTGGGATGTGACAGAGAGAAAGCAGCTCCGGGSCCAAGCTCCAGTGTAAAGA 480  
Db |||||  
Qy 1352 CTTCAAGTGGTCTTGGAGACTGTGTATCCAGAACTGATGTGCTCAGGACAGGCTCGG 1411  
Db |||||  
Qy 481 CTTCAAGTGGTCTTGGAGACTGTGTATCCAGAACTGATGTGCTCAGGACAGGCTCGG 540  
Db |||||  
Qy 1412 CTTCTCGGATGCTCCAGAACTGATGTGTATCCAGAACTGATGTGCTCAGGACAGGCTCGG 1471  
Db |||||  
Qy 541 CTTCTCGGATGCTCCAGAACTGATGTGTATCCAGAACTGATGTGCTCAGGACAGGCTCGG 600  
Db |||||  
Qy 1472 CGATGAAACACAGATGTGGGACACCAGGTCATCTGTATCCCTGTGTATGGATGGGCCA 1531  
Db |||||  
Qy 601 TGATGAAACACAG----- 613  
Qy 1532 GAATCAGTCTTTCAGGTACAGTCCAGAAAGAAATAGCTATTAACACCCACAGGCTGA 1591  
Db |||||  
Qy 614 ----- 613  
Qy 1592 GGGCTGATGCTGTGGAAGCAGGAATGGATACCTTTATCATGCTATCTCTCGGAAGAAC 1651  
Db |||||  
Qy 614 -----ATCATGCTATCTCTCGGAAGAAC 636  
Qy 1652 TGCCCCAGAGAAATCAGAACTTCACTTTGCGAGGAGATGGATCTTTATTTACGACACAGTC 1711  
Db |||||  
Qy 637 TGCCCCAGAGAAATCAGAACTTCACTTTGCGAGGAGATGGTCTTTATTTACGACACAGTC 696  
Qy 1712 CAAGAAATGTGTGAGGCTGGAGGAGGAGTCAAGTGTGACAGTTTGTCTTCACTTCTTACG 1771  
Db |||||  
Qy 697 CAAGAAATGTGTGAGGCTGGAGGAGGAGTCAAGTGTGACAGTTTGTCTTCACTTCTTACG 756  
Qy 1772 AGACTGCACCAACTCGGATCATCAGAAATGCTTCTTCAAGAGCGCATGTTATGAGCCT 1831  
Db |||||  
Qy 757 AGACTGCACCAACTCGGATCATCAGAAATGCTTCTTCAAGAGCGCATGTTATGAGCCT 816  
Qy 1832 CGTGTATCAAGAGGCCATCGAAGGAGACTGTGAGGAGGAGTCTGCCCCAACAAAGACT 1891  
Db |||||  
Qy 817 TGTGTATCAAGAGGCCATCGTAGGAGACTGTGAGGAGGAGTCTGCCCCAACAAAGACT 876  
Qy 1892 TAGCTAAGCAGTACAGAACCCACCAAACTAGGCTGATGCTTTTGAAGGCGCAATC 1951  
Db |||||  
Qy 877 TAGCTAAGCAGTACAGAACCCCTCCAAAACCTAGGCTGCTCTCTTTGAGGAGGGAATC 936  
Qy 1952 ATTTTGCATTTGTGAAAGTTGTGATTTAGTAAATGTAATGAACTTTGTACTT 2011  
Db |||||  
Qy 937 ATTTTGCATTTGTGAAAGTTGTGATTTAGTAAATGTAATGAACTTTGTACTT 996  
Qy 2012 ATTTTGAGAACTTTTAAATGTTCCAAAATACCTATTTTCAAGGGTAAATCGTAAGATG 2071  
Db |||||  
Qy 997 ACTTTGAAAATTTTAAATGTTCCAAAATACCTATTTTCAAGGGTAAATCATAGATG 1056  
Qy 2072 TTAACCTTGGTATTTAGAAAATTTAAACCTTATAATTTTCTATCAARWRWAWATT 2131  
Db |||||  
Qy 1057 TTAACCTTGGTATTT-AAAAATTTAAACCTTATAATTTTCTATCAAGATGTATATT 1115  
Qy 2132 TTACAGTGGTCTTTTACTCTCATTTAGCAAAAAGATAAGATTTTATTTGGTATTTA 2191  
Db |||||  
Qy 1116 TTACAGTCAAGCCCTTTTACTTTTCAATAGC--AAAAGTAAGATTTTATTTGGATATTTA 1173  
Qy 2192 CAAGAATTTCCAGGTACGAAGATATCTGCATGGGTGAAATTCAGGTTCAAGCAAGTACT 2251  
Db |||||  
Qy 1174 CAAGAATTTCCAGGTACGAAGATATCTGCATGGGTGAAATTCAGGTTCAAGCAAGTACT 1233  
Qy 2252 TTGATTAATCTGATTAATCCTCAGCTCGGGGTTTAAAGTTTTCACGATATAGAGACTG 2311  
Db |||||  
Qy 1234 TTGATTAATCTGATTAATCCTCAGCTCGGGGTTTAAAGTTTTCACGATATAGAGACTG 1293  
Qy 2312 TCACTAGGAATGTGATTTGATTTTATTTTCAAGTCAATGAGATCTTCTAGATGATTTTAA 2371  
Db |||||  
Qy 1294 ACACAGGAACACTGTTATGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAA 1353

Qy 2372 AAGAATGCTTTTGGTATGTTGCTTACACAGTTAAACACTCCATATGTTTCATGTCAG 2431  
Db |||||  
Qy 1354 AAGAATGCTTTTGGTATGTTGCTTACACAGTTAAACACTCCATATGTTTCATGTCAG 1413  
Db |||||  
Qy 2432 CCAAAGAGGACTAACCAAGCTGAATCTCAGAGAACAAATTTGCTTTACTAAGCTGAGTC 2491  
Db |||||  
Qy 1414 CCAAAGAGGACTAACCAAGCTGAATCTCAGAGAACAAATTTGCTTTACTAAGCTAAGTA 1473  
Db |||||  
Qy 2492 AACTTGAGAGC-GAACTTCTAAAGTCCGCACTGTAGTGTGGCTGGT----- 2539  
Db |||||  
Qy 1474 AACTTGAGAGGACTAACCAAGCTTCTAAGTCCCTCAGTATGTTGGTGGTGGCTTTAGCAA 1533  
Db |||||  
Qy 2540 CTACCACTATGACTTTTAAACATGTTTATATCATTTTAAATTTTATGATACGTTAGTGT 2599  
Db |||||  
Qy 1534 CTACCACTATGACTTTTAAACATGTTTATATCATTTTAAATTTTATGATACGTTAGTGT 1593  
Db |||||  
Qy 2600 CAGGAGAAATGTAATGTTCTATATGAAATCTTTTCAAGTGTGTTTCAATTAACAG 2659  
Db |||||  
Qy 1594 CAGGAGAAATGTAATGTTCTATATGAAATCTTTTACGAGTTTGTTCATTAACAT 1653  
Qy 2660 TTATTAATTTAAATCAGCGTTAGAGTTTGTGCTGCAACTGCTGTGAAATTTCTCTG 2719  
Db |||||  
Qy 1654 TTAGTAATTTAAATCAGCGTTGAGTTTGTGCTGCCCACTGCTGTGAAATTTCTCTG 1713  
Qy 2720 AGTAATTTCTGATTTGTAATGATCCAGACCAACCTTGAGATTTTGTCAACCTGATTAAG 2779  
Db |||||  
Qy 1714 AGTAATTTCTGATTTGTAATGATCCAGACCAACCTTGAGATTTTGTAAACCTGATTAAG 1773  
Qy 2780 TCAATATGATGATTTAAAGATGTGAGAAC 2811  
Db |||||  
Qy 1774 TCAATATGATGATTTAAAGATGTGAGAAC 1805

RESULT 8  
AC007800/c  
LOCUS  
DEFINITION  
Homo sapiens chromosome 9 clone RP11-105G5 map 9, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, 5 unordered pieces.  
AC007800  
AC007800.4 GI:8844153  
VERSION  
HTG; HTGS\_PHASE1.  
KEYWORDS  
Homo sapiens  
SOURCE  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 52737)  
Birren, B., Linton, L., Nussbaum, C., Lander, E., Allen, N., Anderson, M.,  
Baker, J., Baldwin, J., Barna, N., Beckerly, R., Benn, J., Brown, A.,  
Castle, J., Cerny, J., Colangelo, M., Collins, S., Collymore, A.,  
Cooke, P., DeArillano, K., Depayre, E., Devon, K., Dewar, K.,  
Donelan, L., Doyle, M., Ferreira, P., FitzHugh, W., Forrest, C.,  
Funke, R., Gage, D., Galagan, J., Gardyna, S., Gilbert, D., Grant, G.,  
Hagos, B., Heaford, A., Horton, L., Howland, J. C., Jones, C., Kann, L.,  
Karas, A., Lechoczy, J., Lieu, C., Locke, K., Macdonald, P.,  
Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J.,  
Meldrim, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J.,  
Naylor, J., Niloff, M., O'Connor, T., O'Donnell, P., Pavlin, B.,  
Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P.,  
Strange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A.,  
Tesfaye, S., Toruella-Miller, I., Vassiliev, H., Vo, A., Wagner, A.,  
Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.  
Direct Submission  
Submitted (12-JUN-1999) Whitehead Institute/MIT Center for Genome  
Research, 370 Charles Street, Cambridge, MA 02141, USA  
On Jun 30, 2000 this sequence version replaced gi:8705145.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIGR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L871  
Center clone name: 105 G 5

Query Match	36.5%;	Score 1039;	DB 2;	Length 52737;
Best Local Similarity	96.8%;	Pred. No. 1.2e-159;		
Matches 1139;	Conservative 5;	Mismatches 25;	Indels 8;	Gaps 8;
QY	1682	GGAGGATGGATCTTTATTTTACGAAACAGTGCACGAAGAAATGTGTCCAGGCTGCGAGGAAGGA	1741	
DB	24939	GTAGGATGGATCTTTATTTTACGAAACAGTCCAGAAATGTGTCCAGGCTGCGAGGAAGGA	24880	
QY	1742	GTGAGTGACAGTTTCGTTTCCATCTTTACGAGACTGCAACCAATCGGATCATCAGAAATG	1801	
DB	24879	GTCCAGTGACAGTTTCGTTTCCATCTTTACGAGACTGCAACCAATCGGATCATCAGAAATG	24820	
QY	1802	GTTCTTTCAAAGAGCGCATGTTATGAAGCCTCGTGTATCAAGGAGGCCATCGAAGGAGACT	1861	
DB	24819	GTTCTTTCAAAGAGCGCATGTTATGAAGCCTCGTGTATCAAGGAGGCCATCGAAGGAGACT	24760	
QY	1862	GTGGAGCCAGGACTCTGCCCAACAAAGACTTAGCTTAAGCAGTGAAC-AGAAACCCACCAAA	1920	
DB	24759	GTGGAGCCAGGACTCTGCCCAACAAAGACTTAGCTTAAGCAGTGAACGACGACGACCAACCAAA	24700	
QY	1921	AACCTAGGCTGCATGTCGTTTGAAGAGCGCATCATTTTGCATTTGTGAAGTTGTGTGGA	1980	
DB	24699	AACCTAGGCTGCATGTCGTTTGAAGAGCGCATCATTTTGCATTTGTGAAGTTGTGTGGA	24640	
QY	1981	TTTTAGTAAAAATGTGAATAAGCTTTGTACTTTATTTTGAAGACTTTTTAAATGTTTCCAAAA	2040	
DB	24639	TTTTAGTAAAAATGTGAATAAGCTTTGTACTTTATTTTGAAGACTTTTTAAATGTTTCCAAAA	24580	
QY	2041	TACCCCTATTTTCAAAGGGTAATCGTGAAGATGTTAACCTTGTGA-TTTTAGAATAATTAAA	2099	
DB	24579	TACCCCTATTTTCAAAGGGTAATCGTGAAGATGTTAACCCCTTGTGTANTTTAGAAAAATTAAA	24520	
QY	2100	CCCTATAATAATTTTTCTATCAAAAMRWAAWATTTT-ACAGTCGTGCCTTTTTACTCTCATTA	2158	
DB	24519	CCCTATAATAATTTTTCTATCAAGATGTAATTTTINACAGTCTGCTGCTTTTACTCTCATTA	24460	

Qy	2159	GC	AAAAAGAT	AAAGATTTATTTGGTATTTTAC	-AGAATCCAGGTACG-AGATAT	2216
Db	24459	GC	AAAAAGAT	AAAGATTTATTTTGGTATTTTAC	AGAAATCCAGGTACGNAAGATAT	24400
Qy	2217	CT	GCAATGGT	GGAATCAGGTTCAAGCAA	-CGTACTTTGGCATTAACATGAATACCTCAG	2275
Db	24399	CT	GCAATGGT	GGAATCAGGTTCAAGCAACG	TACTTTGGCATTAACATGAATACCTCAG	24340
Qy	2276	CT	CGGGGTTAAAGTTTTC	-CCAGTATAGAGAGCTGTCAC	TAGGAAACATTTGTATTTGAT	2334
Db	24339	CT	CGGGGTTAAAGTTTTC	CNCAGTATAGAGAGCTGTCAC	TAGGAAACATTTGTATTTGAT	24280
Qy	2335	TAT	TACAGTCA	TGTAGATCTTCTAGATG	-TATTTTAAAGAAATGCTTTTTGGTTATG	2393
Db	24279	TAT	TACAGTCA	TGTAGATCTTCTAGATG	TATTTTAAAGAAATGCTTTTTGGTTATG	24220
Qy	2394	TT	GCTACCA	CAGTTAACCTCCATTAATGTT	CATGTGCAGCCAAAGAGGACTAACCAAGCT	2453
Db	24219	TT	GCTACCA	CAGTTAACCTCCATTAATGTT	CATGTGCAGCCAAAGAGGACTAACCAAGCT	24160
Qy	2454	GAA	ATCTCAG	AGAAACAAATTTTGCTTTACT	TAAGCTGAGTCAACTTGAGAGCGAACTCTTAAC	2513
Db	24159	GAA	ATCTCAG	AGAAACAAATTTTGCTTTACT	TAAGCTGAGTCAACTTGAGAGCGAACTCTTAAC	24100
Qy	2514	AAT	GCGCACTG	TAGTGTGGCTGGTTCTAC	CACTATGACTTTTAAAAACATGTTTATATCAT	2573
Db	24099	AAT	GCGCACTG	TAGTGTGGCTGGTTCTAC	CACATGACTTTTAAAAACATGTTTATATCAT	24040
Qy	2574	TTTT	TAATTTT	TATGATACGGTATGTCAG	GGGAGAAATGTAATGTTCTATATGAATTCCT	2633
Db	24039	TTTT	TAATTTT	TATGATACGGTATGTCAG	GGGAGAAATGTAATGTTCTATATGAATTCCT	23980
Qy	2634	TTTT	CAAGTTCT	GCATTAATAACAGTTATTAAT	TTTAAATCAGCGTTAGAGTTTGTGCTG	2693
Db	23979	TTTT	CAAGTTCT	GCATTAATAACAGTTATTAAT	TTTAAATCAGCGTTAGAGTTTGTGCTG	23920
Qy	2694	CT	GCACTCTG	TGAAAAATTTCTCTGAGTAA	TTCTGATTTTGAATGATCCAGACCAAC	2753
Db	23919	CT	GCACTCTG	TGAAAAATTTCTCTGAGTAA	TTCTGATTTTGAATGATCCAGACCAAC	23860
Qy	2754	CT	GAGATTTT	CTCAACCTGATTAACTCAAT	TATGAATGATTTAAAGAGATGTGAGAACAA	2813
Db	23859	CT	GAGATTTT	CTCAACCTGATTAACTCAAT	TATGAATGATTTAAAGAGATGTGAGAACACT	23800
Qy	2814	AAAA	AAAAAAAA	AAAAAAAAAAAAAAAAAAAAA	AAAAA 2850	
Db	23799	GACT	GCAGATCT	TTTTTGTACTGTAACCA	AAAAATAAA 23763	

Fragment Name	Begin	End	LOCUS AC132018	Accession AC132018
AC132018_0	1	110000		
AC132018_1	100001	210000		
AC132018_2	200001	310000		
AC132018_3	300001	410000		
AC132018_4	400001	463632		
Continuation (5 of 5) of AC132018 from base 400001 (AC132018.Rattus				

**FEATURES**  
**SOURCE**

QY	588	GAAGTGATCCTTGTAGATGACTACAGTATGATAGAGACACCTGAAGGAGCGCCTTGGCCAAT	647
Db	44246	GAGGTCAATTCCTGTGTAGATGACTACAGCGACAGAGATCACCTGAAGGAGCGCCTTGGTCAA	44304
QY	648	GAGCTTTCGGGACTTGCCTCAAGGTGCGCTCATCCGCGCCCAACAGAGAGAGGCGCTGGTG	707
Db	44305	-GAGCTTTCAGAGTGCCTTCAAGGTGCGCTCATCCGCGCCCAACAGAGAGAGGCGCTGGTG	44361
QY	708	CGAGCCCGGCTGTGGGGGGCTGTGGCGGAGGCGGCGATGTTCTGACCTTCCCTGGACTGT	767
Db	44362	CAAG-CCGGCTACTGGGAGCCTTTCGGCGCCAGGGCGAAGTGTCTGAGTTTCTGGACTGT	44420
QY	768	CACGTGTAGTCCACGAAGGGTGGCTGGAGCCGCTGCTGCAGAGGATCCATGAAGAGGAG	827
Db	44421	CACGTGTAGTGTATGAGGGGTGGCTGGAGTCCCTGCTGCAGAGGATCCACGAGAGGAG	44480
QY	828	TCGGCAGTGTGTGCCCGGTGATTGATCGATCGACTGGGAACACCTTCGATACCTGGGG	887
Db	44481	TCGGCGTGTGTGCCCGGTGATTGATCGATCGACTGGGAACACCTTCGATACCTGGGG	44540
QY	888	AACTCCGGGAGGCCAGATCGCGGTTTCGACTGGAGGCTGGTGTTCAGCTGGCACACA	947
Db	44541	AACTCCGGAGAGCCCCAGATTGGTGGCTTGACTGGCGCCTAGTATTCAGCTGGCAGGTG	44600
QY	948	GTTCTGTAGAGGAGAGGATACCGATGCAATCCCGCGTGCATGTCATCAGGTCTCCAACA	1007
Db	44601	GTCCCTTGGCGGAGCGGAAGTTCATGCGGACTCCTATTGACATATCAGGTCTCCAAC	44660
QY	1008	ATGGCTGTGGGCTGTGTTGCTGTGAGTAAAGAAATATTTGAATATCTGGGCTCTATGAT	1067
Db	44661	ATGGCTGTGGGACTGTGTTGTTGTGAGTAAAGAAATATTTGAATATCTGGGCTCTATGAT	44720
QY	1068	ACAGGAATGGAAGTTTGGGAGAGAGAAACCTCGAATTTTCTTTAGGATCTGGCAGTGT	1127
Db	44721	ACCGAATGGAAGTTTGGGAGAGAGAAACCTTGAATTTCTCTTTAGGATCTGGCAGTGT	44780
QY	1128	GGTGGGGTCTGGAACACACACCATGTTCCATGTTGGCCATGTTTTCCTCAAGCAAGCT	1187
Db	44781	GGT- -GACTCTGGAACACACACCCTTACTCCACGTTGGCCACGCTTCCCTTAAGCAAGCT	44838
QY	1188	CCCTACTCCGCAACAGGCTCTGCGCAACAGTGTTCGTCAGCTGATGATGATGAT	1247
Db	44839	CCCTATTCAGCAGAGGCTCTAGCCACAGTGTCCGAGCTCGAAGAGTGTGGACGGAT	44898
QY	1248	GAAATTAAGAGCTCTACTACCATCGCAACCCCGCTGGCCGCTTGAACCTTTTGGGAT	1307
Db	44899	GAACTTAAGAACTCTACCAACCCCGCAACCCCGCTGGAAACCTTCTGGGAG	44957
QY	1308	GTGACAGAGAGGAGCTCCGGGCAAGCTCCAGTGTAAAGACTTCAAGTGGTTCTTG	1367
Db	44958	CTGACAGAGAGAAAGAGCTTCAGCCCAAGCTCCAGCTCGAAGAGTGTGGACGGAT	45017
QY	1368	GAGACTGTGTATCCAGACTGCTGCTGAGGACAGGCTGCTTCTCGGAGTGTCTC	1427
Db	45018	GCTACTGTATCCAGAACTGCTGCTGAGGACAGGCTGCTTCTCGGGAAGCTT	45077
QY	1428	CAGAACAAAGGACTAAGAGACTACTGCTTGAATATAACCTCCGATGAAACACAGATT	1487
Db	45078	GAGAACAGAGGACTACGGGGTACTGCTGGACTATATCTCTCCAGTGAACAAATGTG	45137
QY	1488	GTGGGACACAGGTCTATCTGTACCTCTGTGATGGATGGCCAGAAATCAGTTTTTTCGAG	1547
Db	45138	GAAGGCCACAGGTCTCTTGTACCTCTGCGACGGATGGGTGAGAACCAAGTTTTTCGAG	45197
QY	1548	TACAGCTCCAGAAAGAAATACGCTATACACCCACAGGCTGAGGGCTGATGCTGTG	1607
Db	45198	TGTACATCCCGGAGAAATACGGTATAACCCCGCAGCCCGAGGTCTGCTATGAGGTG	45257
QY	1608	GAGCAGGAATGATACCTTATCATGCTATCTGCGAAGAACTGCCCGAGAGATCAG	1667
Db	45258	GAGGAAGGGAAGGAT- -CCTTATCCGGATCTCTGGAAGAGATACCGTTCAGAGAAATCAA	45315
QY	1668	AAGTTCAATCTTGCGAGGAGATGGATCTTTATTTTCAAGAACAGTCCCAAGAAATGTGCCAG	1727
Db	45316	GAGTTCAATCTTGCGAGGAGATGGATCTTTATTTTCAAGAACAGTCCCAAGAAATGTGCCAG	45375
QY	1728	GCTGCCAGGAGAGGAGTCCGAGTGCATGTTTCTTCAAGAGCGCATGTTATGAAGCCTCGTGTATCAAGGAGCC	1847
Db	45376	GCACAGAGGAGGAGTTCACACTCTGCTTTCGCTCCGAACCTTGCGAGACTGTAGCAACTCA	45435
QY	1788	GATCATCAGAAATGTTTCTTCAAGAGCGCATGTTATGAAGCCTCGTGTATCAAGGAGCC	1847
Db	45436	GACAGCCAGAGGTGGTCTTCAAGGAGCAGATGTCATAGTGCACGGCTTGTGAGGAG	45495
QY	1848	CATCGAAGAGAGACTGTGGAGCCAGGACTCTGCCCAACAAAGACTTAGCTAAGCAGTGAAC	1907
Db	45496	GAGCCACCGCAGCTGTGGCTGTGTGCC- - - - -	45524
QY	1908	AGAACCCACCAAAACCTAGCTGCTTCTTGAAGAGGCAATCATTTTGCCTATTTGTGA	1967
Db	45525	- - - - -CAGTGAGAGGAGCAAAAGTACGTGTTGTCTATCTGTGA	45561
QY	1968	AAGTTGTGTGGATTTAGTAAATGTAATAGCTTTTGTACTTATTTTGGAGAACTTTTT	2027
Db	45562	A- - -TGTGTGGACTCCACACAACTGTGAATATATTTGTACAGATTTGGAAAACCTTTAA	45618
QY	2028	AAATGTTCCAAATACCTTATTTTCAAGGGTAAATCGTAAGATGTTAAACCTTGGTATTT	2087
Db	45619	AAATGTTTCCAGTATCCGTTTCTTAAGGGCGTCTTAAGATGCTGCTTGTGTTTGTG	45678
QY	2088	AGAAATTTAAACCTTATATATTTTCTATCAABAWAWATTTTACAGTCTGCTGCTTT	2147
Db	45679	GGAACTTTAATGACGA- - - - -ACTGCATTTTACCCATGCTGCTTT	45720
QY	2148	TACTCTCATTAGCAAAAAAGATAAGATTTTATTTTGGTATTTTACAGAAATTTCCAGGTA	2207
Db	45721	TATTTCTCATTAGCAGAAAGGGAATATTTTACTTTTGTACTTACAGAGTCTCCGGGTG	45780
QY	2208	CGAAGATATCTGCATGGGTGGAAATCAGGTTCAAGCAACGTAATTTTGCATTAACGATAA	2267
Db	45781	TTTAGTATATCTTGA- - - - -CACAACTGGTAAATACCTCAAAATGTCGACTGTAA	45829
QY	2268	TACCTCA- -GCTGCGGGGTAAAGTTTTCAGTATAGAGAGCTGTACTAGAACATTT	2325
Db	45830	TACCTCAATGTATCAAGTTAGAGTTCTCTCAGGGGATGAGATGGACACTAGAAAATCT	45889
QY	2326	GTATTCATTTATTTTCAG- - - - -GTCAATTCAGATCTTCTAGATGTAATTTTAAAGAAATGC	2379
Db	45890	GTCTTAATTTTGGTGTGGTGGGAGCAGTCTAGATTTTACAGAAAGAAATGC	45949
QY	2380	TTTTGGTTTATGTTGTCTACCAAGTTAAACACTC- - -CATATGTTCTATGTGAGGCAAA	2436
Db	45950	TTTTTGTCTTATGCTGTGATGCCACAGTCAAACTCCATCATACGCTCAGACTGACCCAA	46009
QY	2437	GAGGACTAACCAAGCTGAAATCTCAGAG- - -AACAAATTTGCTTTTACTAAGTGTCAACT	2495
Db	46010	ACAGGCTCACCATCTCTGAAAGCTCAAGAAAGAAATATGTTTGGCAAGGCTCTACAAAC	46069
QY	2496	T- - -GAGAGCGAACTTCTAACAATGCGCACTGTAGTGTGGCTGCT- - - - -T	2539
Db	46070	TTGCGAGTGTGATTTTCAGAGAAATGCTAGTGGAGGATGCTAGTGTACTTAGGAGACT	46129
QY	2540	CTACCACTATGATTTTAAACAATGTTTATATATCATTTTAAATTTTT- - - - -ATGATACGGTA	2595
Db	46130	TAACTACTGTGACTGGAAAAGCAATTTATATCATTTCTAGTTTAAATCAACAAATGACA	46189
QY	2596	GTGTGAGGAGAAATGTAATGTTCTATATGAATTTCTTTTCAAGTTTGTTCATTAATA	2655
Db	46190	GTTCAGGAGAAACGTAATGTTCTACATGAACCTCAACTTTTAAAGTTTTCCTTAATA	46249
QY	2656	ACAGTTTATTAATTTAA	2671
Db	46250	CTATTTGATTTTTTAA	46265



## RESULT 10

AC098559.2/c

## WPCOMMENT

Sequence split into 8 fragments LOCUS AC098559 Accession AC098559

Fragment Name

Begin End

AC098559\_1 1 110000

AC098559\_2 100001 210000

AC098559\_3 200001 310000

AC098559\_4 300001 410000

AC098559\_5 400001 510000

AC098559\_6 500001 610000

AC098559\_7 600001 710000

AC098559\_8 700001 794616

Continuation (3 of 8) of AC098559 from base 200001 (AC098559 Rattus norvegicus clone CH2)

Query Match 32.3%; Score 921.6; DB 2; Length 110000;

Best Local Similarity 69.4%; Pred. No. 2.1e-140;

Matches 1552; Conservative 1; Mismatches 555; Indels 128; Gaps 17;

Qy	468	TATGATTATGATAATTTGCCAGGACATCTGTTATCATAGCATTTTATATGAAGCTGG	527
Db	95301	TATGATTATGATAACCTGCCCAAGACATCCGTTGTCTATGATATTTATATGAAGCTGG	95242
Qy	528	TCAACTCTCTTCGGACAGTTTACAGTGTCTTTCAGACATCCCGGATATCCTGCTAGAA	587
Db	95241	TCCACACTCTTCAGACAGTTTACAGTGTCTTCAGACTTCCCTGTATATCCTGCTGGAG	95182
Qy	588	GAAGTGATCTCTTGTAGATGACTACAGTGTAGAGAGACCTGAGGAGCGCTTGGCGAAT	647
Db	95181	GAGGTCACTTCGTGTAGATGACTACAGGACAGAGATCACCTGAGGAGCGCTGGTCAA	95123
Qy	648	GAGCTTTCCGACATGCCCAAGTCCGCTGATCCGCCCAACAGAGAGAGCGCTGGTG	707
Db	95122	-GAGCTTTCACAGCTGCCCTAGTGTGACCTGATCCCTGCTAGCA--AGAGAGACCTAGTG	95066
Qy	708	CGAGCCCGGTGCTGGGGGGCTGTCGGCGAGGGGGGATGTTCTGACCTTCTCGGACTGT	767
Db	95065	CAAG-CCGGCTACTGGGAGCTTTCGGCCAGGGCCGAGTGTCTGAGTTTCTGGACTGT	95007
Qy	768	CACGTGAGTGCCACGAGGGTGGCTGGAGCGCTGTCGAGAGGATCCATGAAGAGGAG	827
Db	95006	CACGTGACTGTCTAGGGGGTGGCTGGAGTCCCTGCTGAGAGGATCCACGAGAGGAG	94947
Qy	828	TCGSCAGTGTGTCCTCGGTGATGTGATGATCACTGGACACCTTCGATACCTCGGG	887
Db	94946	TCGSCCGTGTGTGCCCGGTATCATGATGTCTGACTGGAGACACCTCTGAGTACCTGGC	94887
Qy	888	AACCTCCGGGAGCCCAAGATCGGCGGTTTCGACTGGAGGCTGGTTCAGCTGGCACACA	947
Db	94886	AACCTCCGGAGAGCCCAAGATGGTGGCTTTCGACTGGCGCTAGTATTCAGCTGGCAGTG	94827
Qy	948	GTTCTGAGAGGAGAGGATACGATGCAATCCCGCTCGATGTCAGTCTCCAAACA	1007
Db	94826	GTCCCTTGGCGGAGCGGAGTGTGCGGACTCCTATTCAGATATTCAGGTCTCCAACT	94767
Qy	1008	ATGCTGTGGCTGTGTTGCTGAGTAAGAAATTTTGAATATCTGGGCTCTTATGAT	1067
Db	94766	ATGCTGTGGGAGCTGTGTTGTTGTGAGTAAGAGATATTTGAATACCTTGGGCTCTTATGAT	94707
Qy	1068	ACAGGATGGAAGTTTGGGAGGAGAAACCTCGAAATTTCTTTAGGATCTGGCAGTGT	1127
Db	94706	ACCGGATGGAAGTCTGGGAGGAGAAACCTTGAATCTCTTTAGGATCTGGCAGTGT	94647
Qy	1128	GGTGGGTTCTGGAACACACCCATGTTCCCATGTTGGCCATGTTTCCCAAGCAAGCT	1187
Db	94646	GGT--GACTCTGGAACACACCCCTTACTCCCACTGGGCGCACGCTCTTCCCTTAAGCAAGCT	94589
Qy	1188	CCCTACTCCCAACAGGCTCTGGCCAAACAGTGTTCGTGACGCTGAGTATGATGGAT	1247
Db	94588	CCCTATTACGACAGAGGCTCTAGCCAAACAGTGTTCGAGCTGAGAGTGTGACCGAT	94529
Qy	1248	GAATTTAAGAGCTCTACTACCATCGCAACCCCGCTGCGCCCTTGGGAACCTTTTGGGGAT	1307

Db	94528	GAACCTTAAAGAACTCTACCAACCCCGGAAACCCCC--AGCCCCCTTGGAACCCCTCTGGGGAG	94470
Qy	1308	GTGACAGAGAGAGAGCAGCTCCGGGACAGCTCCAGTGTAAAGACTTCAAGTGGTCTTTG	1367
Db	94469	CTGACAGAGAGAGAGAGCTTCAGCCAGCTCCAGTGTGAAGACTTAAAGCGGTCTCTG	94410
Qy	1368	GAGACTGTGTATCCAGAACTGCATGTGCTGAGGACAGCCCTGGCTTCTTCGGGATGCTC	1427
Db	94409	GCTACACTGTATCCAGAACTGCATGTGCTGAGGACAGCCCTGGTCTTCTTCGGGAGCTT	94350
Qy	1428	CAGAAACAAAGGACTAACAGACTTACCTTTGATATATAACCTCCCGATGAACACAGATT	1487
Db	94349	GAGAACAGAGGACTACGGGGTACTCCCTGGACTATAATCTCCCACTGAAACAAATGTG	94290
Qy	1488	GTGGGACACAGCTCACTCTGCTACCTCTGATGGATGGCCAGAACTCAGTTTTTCGAG	1547
Db	94289	GAAGGCCACAGCTCTTTGTACCTCTGCCAGGGATGGGTGAGAACAGATTTTTCGAG	94230
Qy	1548	TACACGTCCAGAAAGAAATAGCTATACACCCACCCAGCTGAGGGCTGCAATTTGCTGTG	1607
Db	94229	TGTACATCCCGCAAGAAATACGGTATACACCCCGCCAGCCCGCTGTCATAGCAGTG	94170
Qy	1608	GAAGCAGGAATGATACCTTATCATGCTCTCTGCGAAGAACTGCCCCAGAGATCAG	1667
Db	94169	GAGGAAGGGAAGGAT--CCTTATCCCGGATCTCTGGAAGAGATACCGTTCCAGAGAAATCA	94112
Qy	1668	AGTTTCATCTTCAGGAGGATGATCTTTATTTTCAGAAACAGTCCCAAGAAATGTTCCAG	1727
Db	94111	GAGTTTCATCTTCAGGAGGAGCGACGTTAGTCTTCAAGCAGAGCGGGAATGTTGGAG	94052
Qy	1728	GCTGCGAGGAAGGAGTGCAGTGCAGCTTTCGTTTCCACTCTTACGAGACTGCACCAACTCG	1787
Db	94051	GCCACAGAGGAGGAGTTCACCTGCTGCTTTCCTCCGAACTTGGAGACTGTAGCAACTCA	93992
Qy	1788	GATCATCAGAAATGTTTCTTCAAGAGCGGATGTTATGAAGCCCTGTTATCAAGAGCC	1847
Db	93991	GACAGCCAGAGGTGTTCTTCAAGGAGCACATCTCATAGTGCACGGCTTGTCCAGGAAG	93932
Qy	1848	CATCGAAGGAGTGTGGAGCCAGGACTCTGCCCAACAAAGACTTAGCTAAGCAGTGACC	1907
Db	93931	GAGCCACCGCAGCTGTGGGCTGTGTGCC-----	93903
Qy	1908	AGAAACCCACCAAAACTAGGCTGCATTTGTTGAAGAGGCAATCATTTTGGCCATTTGTGA	1967
Db	93902	-----CAGTGAGAGGACAAAGTACGTGCTTCTCATCTGTA	93866
Qy	1968	AGTGTGTGTTGATTTAGTAAATGTGAATAAGCTTTGTACTTATTTTGAAGAACTTTT	2027
Db	93865	A---TGTGTGGACTCCACAACTGTGAATTTATTTGTACAGATTTGGAAACTTTAA	93809
Qy	2028	AAATGTTCCAAATACCTTATTTTCAAGGGTAAATCGTAAGATGTTAAACCTTGGTATTT	2087
Db	93808	AAATGTTCTCCAGTATCCCGTTTCTTAAGGGCGCTCTAAGATGCTGCTTGGTATTTG	93749
Qy	2088	AGAAATTTAAACCTTATATATTTTCTATCAARAWAWATTTTACAGTCGTGCTCTT	2147
Db	93748	GGAGCTTATGACGA-----ACTGCATTTTACCCATGTTGCTCTT	93707
Qy	2148	TACTCTATTAGCAAAAGATAAGATTTTATTTTGGTATTTTACAGAAATTTCCAGGTA	2207
Db	93706	TATTTCTATTAGCAAGGGGAAATATTTTACTTTTGTACTTACGAAGCTTCCCGGTG	93647
Qy	2208	CGAGATATCTGATGGTGGTAATCAGGTTTCAAGCAACGTACTTTTCATTTAATCTGATAA	2267
Db	93646	TTTAGTATATCTGA-----CACAACTGGTGAATACCTCAATGTCGACTGGTAA	93598
Qy	2268	TACTCTA--GCTCGGGGTTTAAAGTTTTTCCAGTATATAGAGAGACTGTCTAGGAAACATT	2325
Db	93597	TACTCAATGTCTCAAGTTAGTGTCTCTCAGGGGATGAGATGGACACTAGAAATACT	93538
Qy	2326	GTATTTGATTTTACG-----GTCAATGAGATCTTCTAGATGATTTTAAAAAGATGC	2379
Db	93537	GTCTTAATTTTCGTTTGTGTGAGTGGGTGGGACAGTCTAGATTTTATCAAGAAAGATGC	93478



QY 2380 TTTTGGTTATGTTCTGCTACCAAGTTAAACATCTC---CATATGTTTCATGTGACGCCAAA 2436  
D 93477 TTTTGGTTATGTTCTGCTACCAAGTTAAACATCTC---CATATGTTTCATGTGACGCCAAA 93418  
QY 2437 GAGGACTAACCAAGCTGAATCTCAGAG-AACAATTTGCTTTTACTTAAGCTGAGTCAACT 2495  
D 93417 ACAGGCTACACATCTCTGAAGAGCTCAAGAAAGAAATATGTTTGGCCAAAGTCTCTACAAAC 93358  
QY 2496 T---GAGAGCGAACTTCTTAACTACGCGCACTGTAGTGTGGCTGGT-----T 2539  
D 93357 TTGGAGTGTGATTTTCCAGCAATGCTAGTGGAGGATCTAGTACTTAGGAGGACT 93298  
QY 2540 CTACCACTATGACTTTTAAACATGTTTATATCATTTTAAATTTT-----ATGATACGGTA 2595  
D 93297 TAACTACTGTGACTGGGAAAGCAATTTATATCATTTTCTAGTCTTTTAAATCAACAATGACA 93238  
QY 2596 GTGTGAGGAGAAATGTAATGTTCTATATCAATGCTTCTTTTCAAGTCTGTTTCAATTAATA 2655  
D 93237 GTTCCAGGAGAAACGTAATGTTCTATCAATGCTTCTTTTCAAGTCTGTTTCAATTAATA 93178  
QY 2656 ACAGTTATTAAATTAA 2671  
D 93177 CTATTGATTTTAA 93162

## RESULT 11

AC129139 288118 bp DNA linear HTG 11-OCT-2002  
LOCUS Rattus norvegicus clone CH230-293111, \*\*\* SEQUENCING IN PROGRESS  
DEFINITION \*\*\* 14 unordered pieces.  
AC129139  
VERSION AC129139.2 GI:23829088  
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS\_ENRICHED.  
SOURCE Rattus norvegicus (Norway rat)  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

## REFERENCE

AUTHORS Muzny, D., Marie, J., Metzker, M., Lee, S., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flegel, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gili, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogue, M., Hollins, B., Howells, S., Hui, Y., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuewa, L., Loulseghe, H., Lozano, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwankwelen, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,

Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., L., Puazo, B., Quintero, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D., Sneed, A., Soderstrom, E., Song, X.-Z., Sorelle, R., Sosa, J., Steinhilber, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlarczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstein, G. and Gibbs, R. A.

Unpublished  
2 (bases 1 to 288118)  
Worley, K. C.

Direct Submission  
Submitted (27-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 288118)  
Rat Genome Sequencing Consortium.

Direct Submission  
Submitted (11-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Oct 11, 2002 this sequence version replaced gi:21998925.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: GRGM  
Center clone name: CH230-293111  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 214989 bases at least Q40  
Consensus quality: 222817 bases at least Q30  
Consensus quality: 227320 bases at least Q20  
Estimated insert size: 252471; sum-of-contigs estimation  
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html))  
\* NOTE: This sequence may represent more than one clone.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 14 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 5778: contig of 5778 bp in length  
\* 5779 5878: gap of unknown length  
\* 5879 16373: contig of 10495 bp in length

```
* 16374 16473: gap of unknown length
* 16474 90420: contig of 73947 bp in length
* 90421 90520: gap of unknown length
* 90521 236782: contig of 146262 bp in length
* 236783 236882: gap of unknown length
* 236883 250122: contig of 13240 bp in length
* 250123 250222: gap of unknown length
* 250223 257033: contig of 6811 bp in length
* 257034 257133: gap of unknown length
* 257134 259071: contig of 1938 bp in length
* 259072 259171: gap of unknown length
* 259172 261008: contig of 1837 bp in length
* 261009 261108: gap of unknown length
* 261109 264448: contig of 3340 bp in length
* 264449 264548: gap of unknown length
* 264549 267548: contig of 3000 bp in length
* 267549 271482: contig of 3834 bp in length
* 271483 271582: gap of unknown length
* 271583 275622: contig of 4040 bp in length
* 275623 275722: gap of unknown length
* 275723 286107: contig of 10385 bp in length
* 286108 286207: gap of unknown length
* 286208 288118: contig of 1911 bp in length.
FEATURES
    source
        1..288118
            /organism="Rattus norvegicus"
            /mol_type="genomic DNA"
            /db_xref="taxon:10116"
            /clone="CH230-293111"
        1..1287
            /note="wgs_end_extension"
            clone_end:T7"
        1338..5778
            /note="wgs_end_extension"
            clone_end:T7"
        3086..3988
            /note="clone_boundary"
            clone_end:T7
            site:Mbol
            end_sequence:RXAIQ54TJB"
        5879..7959
            /note="wgs_contig"
            12280..13622
            /note="wgs_contig"
            14019..16373
            /note="wgs_contig"
            16474..19067
            /note="wgs_contig"
            19118..21281
            /note="wgs_contig"
            21332..22964
            /note="wgs_contig"
            28345..31644
            /note="wgs_contig"
            31695..34111
            /note="wgs_contig"
            40505..44913
            /note="wgs_contig"
            44964..47869
            /note="wgs_contig"
            47920..49820
            /note="wgs_contig"
            49871..52672
            /note="wgs_contig"
            52723..56843
            /note="wgs_contig"
            61887..63717
            /note="wgs_contig"
            72165..76565
            /note="wgs_contig"
            78769..86319
            /note="wgs_contig"
```

```
misc_feature 86370..90420
              /note="wgs_contig"
misc_feature 90521..92438
              /note="wgs_contig"
misc_feature 92489..94959
              /note="wgs_contig"
misc_feature 96097..99542
              /note="wgs_contig"
misc_feature 236883..239542

Query Match      32.0%; Score 911.4; DB 2; Length 288118;
Best Local Similarity 68.8%; Pred. No. 1.1e-138;
Matches 1540; Conservative 1; Mismatches 567; Indels 129; Gaps 16;

QY 468 TATGATTATGATAATTTGCCAGGACATCTGTATCATAGCATTTTATATGAAGCTGG 527
DB 83581 TATGATTATGATAAACCCTGCCAAGACATCCGTTGTATAGCATTTTATATGAAGCTGG 83640
QY 528 TCAACTCTCTCTCGGACAGTTTACAGTGTCTTTCAGACATCCCGGATATCCTGTAGAA 587
DB 83641 TCCACACTCTCTCAGACAGTTTACAGTGTCTCGAGACTTCCCTGATATCCTGTGGAG 83700
QY 588 GAAGTGATCTCTGTAGATGACTACAGTATAGAGACACTGAGAGAGCGCTTGGCAAT 647
DB 83701 GAGGTCAATTCTGTAGATGACTACAGCGACAGAGATCACCTGAAGGAGCCCTGTCAA - 83759
QY 648 GAGCTTTCCGGACTGCCAAGGTGCCCTGTATCCGCGCAACAAGAGAGAGGGCCCTGGTG 707
DB 83760 -GAGCTTTCACAGTGCCTTAGGTGCACCTGATCCCTGCTAGCA--AGAGAGGACCTAGTG 83816
QY 708 CGAGCCCGGCTGTGGGGGCGTCTCGGCGAGGGGCGATGTTCTGACCTTCTTGGACTGT 767
DB 83817 CAAG-CCGCTACTTGGAGCGCTTTGGCGCCAGGGCGAGTGTCTGAGTTTCTGGACTGT 83875
QY 768 CACTGTGATGCCACGAAGGGTGGCTGGAGCCGCTCTGAGAGGATTCATGAAGAGGAG 827
DB 83876 CACTGTGACTGTCTATGAGGGTGGCTGGAGTCCCTCTGCAGAGGATCCACGAGAAGGAG 83935
QY 828 TCGGCAGTGGTGGCCCGGTGATTGATGTGATCGACTGGGAACACCTTCGAATACCTGGGG 887
DB 83936 TCGGCCGTGGTGGCCCGGTATCGAATGATTCATGTAACACCTCTGAGTACCTGGGC 83995
QY 888 AACTCCGGGAGGCCCCAGATCGGCGGTTTCGACTGGAGGCTGGTGTTCAGTGGCGACACA 947
DB 83996 AACTCCGGAGAGGCCCCAGATTGGTGGCTTTCGACTGGCGCTAGTATTCAGTGGCAGGTG 84055
QY 948 GTTCTTGAGAGGAGAGGATACGGATGCAATCCCGCGTGCATGTCATCAGGTCTCCACA 1007
DB 84056 GTCCCTTGGCGGAGCGGAAGTTGATCGCGACTCTCTATTGACATTTATCAGGTCTCCAAC 84115
QY 1008 ATGGCTGGTGGGCTGTTGCTGTGAGTAAGAATATTTTGAATATCTGGGGTCTTATGAT 1067
DB 84116 ATGGCTGGGGGACTGTTGTTGTTGAGTAAGAGATATTTTGAATACCTGGGGTCTTATGAT 84175
QY 1068 ACAGGAATGGAAGTTTGGGAGGAGAAAAACCTCGAATTTTCTTTTAGGATCTGGCAGTGT 1127
DB 84176 ACCGGAATGGAAGTCTGGGAGGAGAAAAACCTTGAATTTCTCTTTTAGGATCTGGCAGTGT 84235
QY 1128 GGTGGGGTCTGAAAACACACCATGTTCCCATGTTGGCCATGTTTTCCTCCCAAGCAAGCT 1187
DB 84236 GGT--GACTCTGGAACACACACCCTTACTCCACGTTGGGCCACGCTTTCCTTAAGCAAGCT 84293
QY 1188 CCTACTCCCGCAACAAGGCTCTGGCCACAGTGTTCGTGCAGCTGAAGTATGATGATGAT 1247
DB 84294 CCTATTTCACGACGAGGCTCTAGCCACAGTGTCCGAGCTCAGAAAGTGTGGACCGAT 84353
QY 1248 GAATTTAAAGAGCTCTACTACCATCGCAACCCCGTCCCGCTTGGAAACCTTTTGGGGAT 1307
DB 84354 GAATTTAAAGAACTCTTACCAACCCCGGAAACCCCG--AGCCCGCTTGGAAACCTCTGGGGAG 84412
QY 1308 GTGACAGAGAGGAGAGCAGCTCCCGGCAAGCTCCAGGTAAAGACTTCAAGTGTCTTCTTG 1367
DB 84413 CTGACAGAGAGAAAGAAAGCTTTCAGCCCAAGCTCCAGTGTGAAGACTTTAAAGCGGTTCTCTG 84472
```

Qy	1368	GAGACTGTGATCCAGAACTGCTGTGCTGAGGACAGGCGCTTCTTCGGGATGCTC	1427
Db	84473	GCTACACTGTACCCAGAACTGCTGTGCTGAGGACAGGCGCTTCTTCGGGAGCTT	84532
Qy	1428	CAGAAACAAAGACTAACAGACTACTGCTTTGACTATTAACCTCCCGATGAACACCAT	1487
Db	84533	GAGAAACAGAGACTACCGGGGACTGCTGCTGACTATTAATCTCCAGTGAACAAATGTG	84592
Qy	1488	GTGGGACACCAAGGTCATTCTGTACTCTGTGATGGGATGGGCCAGAAATCAAGTTTTCGAG	1547
Db	84593	GAAGGCCACCAAGGTCCTTTTGTACTCTGTGCAAGGATGGTCAAGAACCAAGTTTTCGAG	84652
Qy	1548	TACAGTCCCAAGAAATACGCTATTAACACCCACAGCCTGAGGGCTGATTCGCTGTG	1607
Db	84653	TGTAATATCCCGGCAAGAAATACGCTATTAACACCCCGCAGCCCAAGGCTGTCATAGCAGTG	84712
Qy	1608	GAAGCAGGAATGGATACCTTTATCATCATCTCTGCGAAGAACTGCGCCAGAGAAATCAG	1667
Db	84713	GAGGAGGGAGGAT--CCTTATCCCGATCTCTGGAAGATACCGTTCCAGAGAAATCAA	84770
Qy	1668	AAGTTACTTTCGAGGAGGATGGATCTTTATTTCCAGAACAGTCCAAAGAAATGTGTCCAG	1727
Db	84771	GAGTTTATCTTACAGGAGGACGACGCTAGTCTTACAGCAGAGCGGGAATGTGTGGAG	84830
Qy	1728	GCTGCGAGGAGGATCGAGTGACAGTTTCGTTCCACTCTTACGAGACTGCACCAACTCG	1787
Db	84831	GCCACAGAGGAGGATTCAACTCTGCTGCTCCGAACTTGCAGAGACTGTAGCAACTCA	84890
Qy	1788	GATCATCAGAAATGGTTCTTCAAGAGCGGATGTTATGAGCCTCTGTATCAAGGAGCC	1847
Db	84891	GACAGCAGAGGTTGCTTCAAGGAGCAGATGTCATAGTGCAGCGCTTGTGCGGAAG	84950
Qy	1848	CATCGAAGGAGACTGTGCGAGCAGGACTCTGCCCAACAAAGACTTAGCTAAGCAGTGACC	1907
Db	84951	GAGCCACCCAGCTGTGGCTGTGTGCC-----	84979
Qy	1908	AGAACCCCAACAAATAGGCTGATGCTTTGAGGAGGCAATCATTTGCCATTTGTGA	1967
Db	84980	-----CAGTGAGAGGACAAAGTACGTGGTGTCTGCTCTGTGA	85016
Qy	1968	AAGTTGTGTTGGATTTAGTAAATGTGAATAGCTTTGACTTATTTTTCAGACTTTT	2027
Db	85017	A-----TGTTGTGGACTCCACAACTGGAATATATTTTACAGATTTGGAACCTTTAA	85073
Qy	2028	AAATGTTCCAAATACCCCTATTTTCAAAGGGTAACTCGTAAGATGTTAAACCTTGGTATTT	2087
Db	85074	AAATGTTCCAGTATCCCGTTTCTAAGGGCGCTCTAAGATGCTGCTTGTGTTATTTG	85133
Qy	2088	AGAAATTAACCTTATATATATTTTCTATCAARAWAWATTTTACGTGCGCTTT	2147
Db	85134	GGAAGCTTAATGCACGA-----ACTGCATTTTACCCATGTTGCGCTTT	85175
Qy	2148	TACTCTCATAGCAAAAGATTAAGATTTTATTTTGGTATTTTACAGAAATTTCCAGGTA	2207
Db	85176	TATTTCTATTAGCAGAAAGGGGAATATTTTACTTTGTTACTTACGAAAGTTCCTCGGGTG	85235
Qy	2208	CGAAGATATCTGATGGTGGAAATACAGTTTCAAGCAACGCTACTTTGCATTAACTGATAA	2267
Db	85236	TTTAGTATACCTGA-----CACAACTGTTATACCTCAATGTCGACTGGTAA	85284
Qy	2268	TACCTCA--GCTCGGGGTTAAAGTTTTCAGTATAGAGAGACTGTCACTAGGAACATT	2325
Db	85285	TACCTCAATGTCAAGTTAGAGTTCTCTCAGGGGATGAGATGGACACTAGAAATACT	85344
Qy	2326	GTATTGATTTTATCAG-----GTCATTGAGATCTTCTAGATGATTTTAAAAAGATGC	2379
Db	85345	GTCTTAATTCGTTGTGTGAGTGGGTGGGACAGTCTAGATTTATACAAGAAAGATGC	85404
Qy	2380	TTTTTGGTTATGTGCTCTACCAAGTTTAACTC-----CATATGTTTCATGTCAGCCAA	2436
Db	85405	TTTTTGTGTTATGCTGATGCCACAGTCAAACTTCCATATACGCTCAGACTGACCCAA	85464
Qy	2437	GAGGACTAACCAAGCTGAAATCTCAGAGAACAAAT-----TGCTTACTAAGCTGAGTCA	2492
Db	85465	ACAGGCTCACCATCTCTGAAAGCTCAAGAGAAATATGTTGTTCCCAAGCTCTAGCAA	85524
Qy	2493	ACTTGAGAGGAACTTCTTAACATCGCGACTGT-----ACTGTGGCTGGT	2538
Db	85525	ACTTGGAGTGATCATTTTCAGCAATGTCTATGGAGGATGTCTAGTACTTAGGAGGAC	85584
Qy	2539	TCTACCACACTACTTCTTAAACATGTTTATATCATTTTAAATTTT-----ATGATACGGT	2594
Db	85585	TTAACTACTGTGACTGMAAAAGCAATTTATATCATTTCTAGTTTAAATCAACAATGAC	85644
Qy	2595	AGTGTGAGGAGAAATGTAATGTCTATATGAAATTCCTTTTCAAGTTTGTTCATTAAT	2654
Db	85645	AGTTCAGGAGAAAGTAAATGTTCTACATGAAACTCAACTTTTAAAGTTTTCCTTAAAT	85704
Qy	2655	AACAGTTTATTAATTTAA	2671
Db	85705	ACTATTGATGTTTAA	85721
RESULT 12			
AC121727			
LOCUS			
DEFINITION			
AC121727			
AC121727.3			
HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.			
Rattus norvegicus (Norway rat)			
Rattus norvegicus			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;			
Rattus.			
1 (bases 1 to 228368)			
REFERENCE			
AUTHORS			
Muzny, D., Marie, E., Metzker, M., Lee, S., Abramson, S., Adams, C., Alder, J.,			
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Angiano, D.,			
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,			
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,			
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,			
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,			
Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A.,			
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,			
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,			
Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,			
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,			
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,			
Egan, A., Escotto, M., Evans, C., Falls, T., Fan, G.,			
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,			
Fraser, C., Gabisai, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,			
Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,			
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,			
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,			
Hernandez, R., Hines, S., Hladun, S., Hodgson, A., Hogue, M.,			
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,			
Jacksom, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,			
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,			
Kowals, C., Kraft, C., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,			
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,			
Lorenshew, L., Loulseghe, H., Lozano, R., Lu, X., Lu, X., Ma, J.,			
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,			
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,			
Mahoney, S., McLeod, M., McNeill, T., Z., Meenen, E.,			
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,			
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,			
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,			
Nwankwelu, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,			
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,			
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.,			
Puro, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M., A., Reigh, R.,			
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,			
Rives, C., Rodkey, T., Rojas, R., Rose, R., Rose, R., Ruiz, S.,			
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,			



Db 163988 TCGCCGCTGGTGTGCCCGTTATCGATGTCAATGACCTGGAAACACCTCTGAGTACCTGGGC 164047  
QY 888 AACTCCGGGAGGCCACAGATCGCGGTTTCGACTCGAGGCTGGTTCAGTGGCCACACA 947  
Db 164048 AACTCCAGAGAGCCCAAGATTGGTGGCTTGACTGGCGCTAGTATTCACGTGCACAGTG 164107  
QY 948 GTTCTCTGAGAGGAGAGATACGGATGCAATCCCGCTGATGTCAATGATGATGATGATGAT 1007  
Db 164108 GTCCCTTGGCGGAGCGGAAAGTGTGATGCGGACTCTATTGACATATATCAGTCTCCAACT 164167  
QY 1008 ATGCTGTGGCTGTGGTGTGGTGTGAGTAAAGATATTTTGAATATCTGGGCTCTATGAT 1067  
Db 164168 ATGCTGTGGGACTGTGTGTGTGAGTAAAGATATTTTGAATATCTGGGCTCTATGAT 164227  
QY 1068 ACAGGAATGGAAGTTTGGGAGGAGAAACCTCGAATTTTCTTTAGGATCTGGCAGTGT 1127  
Db 164228 ACCGGAATGGAAGTTTGGGAGGAGAAACCTTTGAAATCTCTTTAGGATCTGGCAGTGT 164287  
QY 1128 GGTGGGTTCTGAAACACACCAATGTTCCATGTTGGCCATGTTTCCCAAGCAAGCT 1187  
Db 164288 GGT--GACTCTGGAACACACACCTTACTCCACGTTGGGCCACGCTCTTCCCTAAGCAAGCT 164345  
QY 1188 CCTACTCCCGCAACAGGCTCTGGCCACAGTGTTCGTGCAGCTGAAGTATGGATGGAT 1247  
Db 164346 CCTATTACGCAAGAGGCTCTAGCCACAGTGTCCGAGCTGCAGAGTGTGGACGGAT 164405  
QY 1248 GAATTTAAAGAGCTTACTACCATCGCAACCCCGCTGCCGCTTGAACCTTTTGGGGAT 1307  
Db 164406 GAATTTAAAGAACTTACCACACCCCGGAAACCCCGGAGTGTGGGAG 164464  
QY 1308 GTGACAGAGAGAGAGCTCCGGGCAAGCTCAAGCTCAAGTGTAAAGCTTCAAGTGGTCTTG 1367  
Db 164465 CTGACAGAGAGAGAGAGCTTCAAGCCCAAGCTCCAGTGTGAAGACTTAAAGCGGTTCTG 164524  
QY 1368 GAGACTGTGTATCCAGACTGATGCTGCTGAGGACAGGCTGGCTTCTCGGAGTCTC 1427  
Db 164525 GCTCAATGTATCCAGAACTGCAATGTGCTGAGGACAGGCTGGTCTTCTCGGGAAGCTT 164584  
QY 1428 CAGAACAAAGGACTTACAGACTACTGCTTTGACTATAACCTCCCGATGAAACCAAGATT 1487  
Db 164585 GAGAACAGAGAGAGAGGAGTACTGCTGACTATTAATCTCCAGTGAACAGTGTG 164644  
QY 1488 GTGGGACACAGTCAATCTGTACCTCTGTATGGGATGGCCAGAAATCAGATTTTTCGAG 1547  
Db 164645 GAAGGCCACAGGCTCTTTGTATCTCTGACAGAGATGGTCTGAGAAACAGATTTTTCGAG 164704  
QY 1548 TACAGCTCCAGAGAGAGATACGCTATACACCCACAGGCTGAGGCTGATGCTGTG 1607  
Db 164705 TGTACATCCCGGAGAGAAATACGGTATACACCCCGCAGGCTGCTGATAGCAGTG 164764  
QY 1608 GAAGCAGGAATGGATACCTTATCATGCTATCTGCGAGAGAACTGCCCGAGAGATCAG 1667  
Db 164765 GAGGAGGAGAGAT--CTTATTCGGATCTCTGGAAGATACGGTCTCCAGAGAAATCAA 164822  
QY 1668 AAGTTCATCTTCAGGAGGATGGA----TCTTTTATTCAGAAACAGTCCAAAGAAATGTGT 1723  
Db 164823 GAGTTCATCTACAGGAGAGGACAGCAGCTTGTAGTCTCAAGCAGACCGGAAATGTGT 164882  
QY 1724 CAGGCTGCGAGAGAGTCAAGTGTGATGCTTCCACTTACAGAGACTGCAACCAA 1783  
Db 164883 GGAGGCCACAGAGGAGGATCAACTCTGCTTTGTCTCGCACTTCGAGACTGTAGCAA 164942  
QY 1784 CTCGGATCATCAGAAATGGTCTTCAAGAGCGCATGTTATGAAGCTCGTGTATCAAG 1843  
Db 164943 CTCAGACAGCAGAGTGTCTTCGACAGAGCTCATGTATAGTCAGGCGCTTGTGAG 165002  
QY 1844 AGCCCATCGAAGGAGAGCTGTGGAGCCAGGACTCTGCCCAACAAAGACTTAGCTAAGCAGT 1903  
Db 165003 GAAGCAGCCACCGCGCGCGGGGCTGTGTGCC----- 165035  
QY 1904 GACCAGAACCCCAAAACATAGGCTGCTATGCTTTTGAAGAGGCAATCATTTTGGCATTT 1963

Db 165036 -----CAGTGAAGGGGCAAAAGTACGTGGTGTGTCACT 165068  
QY 1964 GTGAAGTGTGTGTGATTTAGTAAATAATCTGAATAAGCTTTTGTACTTATTTTGAAGACT 2023  
Db 165069 GTGAA--TGTGTGGACTCCACAAATGTGAATTTATTTGTACAGATTTGGAAGACT 165125  
QY 2024 TTTTAAATGTTCCAAATAACCTTATTTTCAAGGGTAAATCGTAAAGATGTTTAAACCTTGGT 2083  
Db 165126 TTTAAATATGTTTCCAGTATCCCGTTTCTAAGGGCGCGCTAAGATGCTGCCTTTGGTA 165185  
QY 2084 ATTACAAATTAATAACCTTATAATATTTTCTATCAAAARWAWATTTTACAGTGTGTC 2143  
Db 165186 TTTGGGAGCTTAATGCAAGTGTGCAT-----TTTACCCTATGGTGC 165227  
QY 2144 CTTTACTCTCATTTAGCAAAAAGATAAAGATTTTATTTTGGTATTTTCAAGAATTTCCCA 2203  
Db 165228 CTTTATTTCTCATTTAGCAGAAAGGGGAATATTTTACTTTTGTACTTACGAAGCTTCCCG 165287  
QY 2204 GGTACGAAGATATCTGCATGGGTGGGAATCAGGTTCAAGCAACGTACTTTTGCAATTAACGT 2263  
Db 165288 GGTGTTTGTAGTACTTGA-----CACAAACTGGTAAATACCTCAAAATGTGGACTG 165336  
QY 2264 ATAAATACCTCA--GCTCGCGGGTTAAAGTTTCCCAAGTATAGAGACACTGTCTACTAGGA 2321  
Db 165337 GTAATACCTCAATGTCAATCAAGTTAGGTTCTCTCAGGGGGATGAGATGGACACTAGAA 165396  
QY 2322 CATTGTATTGATTTATTTCAG-----GTCAATTGAGATCTTCTAGATGTATTTTAAAAAGA 2375  
Db 165397 TACTGTCTTAATTTTCAATTTGTTGTTGAGTGGTGGGACAGCTTAGATTTTATCAAGAAGA 165456  
QY 2376 ATGCTTTTGTGTTATGTTGTCTACCAAGTTAAACATC---CATTAATGTTTCAATGTCAGC 2432  
Db 165457 ATGCTTTTGTGTTATGTTGTGATGCCACAGTCAAACTCCATCATTAACGCTCAGACTGAC 165516  
QY 2433 CAAAGAGGACTTAACCAAGCTGAATCTCAGAG--AACATTTGCTTTTACTAAGCTGAGTC 2491  
Db 165517 CCAAACAGGCTCACCCTCTCTGAAAGCTCAAGAAGAAATATGGTTTGCAGGCTCTAC 165576  
QY 2492 AACTT---CAGAGAGCAACTTTTAAACAATGCC-----GCACCTGTAGTGTGGCTG 2536  
Db 165577 AACTTGCAGTGTACATTTTCCAGCAATGCTATGGAGAGTGTCTAGTACTTAGAGG 165636  
QY 2537 GTTCTACCACTATGACTTTTAAACATGTTTATATCATTTTAAATTTT--ATGATACGGTA 2595  
Db 165637 ACTTAACGACTGTGACTGGAAGAAAGCATTCATCATCTCTAGTTTATTAATCAAGACA 165696  
QY 2596 GTGTCAGGAGAAATGTAATGTTCTATATGAATTCCTTTTCAAGTGTGTTCAATTAATA 2655  
Db 165697 GTTCCAGGAGAAACGTAATGTTCTACATGAACCTCAACTTTAAAGTTTTTCCCTAATA 165756  
QY 2656 ACAGTTATTAAATTTAA 2671  
Db 165757 CTATTGATGTTTAA 165772

## RESULT 13

AC129139/c

LOCUS

DEFINITION

AC129139

VERSION

KEYWORDS

SOURCE

ORGANISM

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

REFERENCE

1 (bases 1 to 288118)

Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,

Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,

Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,

Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,

AC129139 288118 bp DNA linear HTG 11-OCT-2002  
Rattus norvegicus clone CH230-293111, \*\*\* SEQUENCING IN PROGRESS  
\*\*\*, 14 unordered pieces.

AC129139.2 GI:23829088

HTG: HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 288118)

Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,

Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,

Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,

Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,



misc_feature	/note="wgs contig" 14019..16373 /note="wgs contig" 16474..19067 /note="wgs contig" 19118..21281 /note="wgs contig" 21332..22964 /note="wgs contig" 28345..31644 /note="wgs contig" 31695..34111 /note="wgs contig" 40505..44913 /note="wgs contig" 44964..47869 /note="wgs contig" 47920..49820 /note="wgs contig" 49871..52672 /note="wgs contig" 52723..56843 /note="wgs contig" 61887..63717 /note="wgs contig" 72165..76565 /note="wgs contig" 78769..86319 /note="wgs contig" 86370..90420 /note="wgs contig" 90521..92438 /note="wgs contig" 92489..94959 /note="wgs contig" 96097..99542 /note="wgs contig" 236883..239542	Db	163292	AATCCAGAGAGCCCGCAGATTGGTGGCTTTGACTGGCGCCTAGTATTACGTGGCAGTG	163233
misc_feature		Qy	948	GTTCCTGAGAGGAGAGATACCGATGCCATCCCGCTCGATGTCACTAGTCTCCCAACA	1007
misc_feature		Db	163232	GTCCCTGGCGGAGCGAAGTTGATGGGACTTCTATTGACATTATCAGGTCTCCAACT	163173
misc_feature		Qy	1008	ATGGCTGGTGGCTGTTTGTGTGAGTAAGAAATATTTTGAATATCTGGGGTCTTATGAT	1067
misc_feature		Db	163172	ATGGCTGGGGGACTGTTTGTGTGAGTAAGAGATATTTTGAATACCTGGGGTCTTATGAT	163113
misc_feature		Qy	1068	ACAGGAATGGAAGTTTGGGAGGAGAAACCTCGAATTTTCCCTTAGGATCTGGCAGTGT	1127
misc_feature		Db	163112	ACCGAATGGAAGTCTGGGAGGAGAAACCTTGAATTTCTCCCTTAGGATCTGGCAGTGT	163053
misc_feature		Qy	1128	GGTGGGGTCTCGAAACACACACCATGTTCCCATGTTGGCCATGTTTCCCAAGCAAGCT	1187
misc_feature		Db	163052	GGT--GACTCAGGAACACACACCTTACTCCACGTGGCCATTTCTCCCTTAGCAAGCT	162995
misc_feature		Qy	1188	CCCTACTCCCGCAACAGGCTCTGGCCAACAGTGTTCGTGCAGCTGAAGTATGATGGAT	1247
misc_feature		Db	162994	CCCTATTACGCAGCAAGGCTCTAGCCAAACAGTGTCCGAGCTGCAGAAAGTGTGCACGAT	162935
misc_feature		Qy	1248	GAATTTAAAGAGCTTACTACCATCGCAACCCCGTCCCGCTTGGAACTTTTGGGGAT	1307
misc_feature		Db	162934	GAATTTAAAGAACTCTACCAACCCCGGAAACCCCG--AGCCCGCTTGGAAACCTCTGGGGAG	162876
misc_feature		Qy	1308	GTGACAGAGAGGAGAGCAGCTCCGGGACAAAGCTCCAGTGTAAAGACTTCAAGTGGTCTTG	1367
misc_feature		Db	162875	CTGACAGAGAGAAAGAAAGCTTACGCCAGCTCCAGTGTGAAAGCTTAAAGCGGTTCTG	162816
misc_feature		Qy	1368	GAGACTGTGTATCCAGAACTGCAATGCTGCTGAGGACAGGCTCGCTTCTTCGGGATGCTC	1427
misc_feature		Db	162815	GCTACAATGTACCCAGAACTGTCATGTGCCAGAGGACAGGCTCGTTCCTTCGGGAACTT	162756
misc_feature		Qy	1428	CAGAACAAAGGACTTACAGACTACTGCTTTGACTATTAACCTCCCGATGAAACCAAGATT	1487
misc_feature		Db	162755	GAGAACAGAGGAGCAGCGGGGTACTGCTGGACTATAATCTCCAGTGAACACAGTGTG	162696
misc_feature		Qy	1488	GTGGGACACCCAGGTCAATCTGTACTCTGTGATGGATGGCGGAGAGTCAAGTCTTCGAG	1547
misc_feature		Db	162695	GAAGGCCACAGGTCTCTTTGTACTCTGCCACAGATGGGTGAGAACAGTCTTCGAG	162636
misc_feature		Qy	1548	TACAGTCTCCAGAAAGAAATACGCTATAACACCCACAGCTCGAGGCTGCATTTGCTGTG	1607
misc_feature		Db	162635	TGTACATCCCGCAGAAATACGGTATAACACCCGACAGCCCGAGGTCTCATAGCAGTG	162576
misc_feature		Qy	1608	GAAGCAGGAATGATACCTTATCATGCTCTCTCGAAGAAACTGCCCCAGAGATCAG	1667
misc_feature		Db	162575	GAGGAAGGGAAGGAT--CCTTATTCGGATCTCTGGAAAGATACGGTTCAGAGAAATCAA	162518
misc_feature		Qy	1668	AGTTTCATCTTGACGAGGATGCA---TCTTTTATTTTCAAGAACAGTCCCAAGAAATGCTG	1723
misc_feature		Db	162517	GAGTTTCATCTACAGAGAGACGACAGCGTTAGTCTACAGCAGACCGGGAATGTGT	162458
misc_feature		Qy	1724	CCAGGCTGCGAGGAAGAGTTCGAGTGCAGTTCCTTTCCTTTCCTTTCAGGAGCTGCACAA	1783
misc_feature		Db	162457	GGAGGCCACAGAGAGGAGTTCACCTCTGCTTCCTTTCCTTTCCTTTCAGGAGCTGTAGCAA	162398
misc_feature		Qy	1784	CTCGATCATCAGAAATGTTCTTCAAAGAGCCATGTTATGAGCCCTCGTGTATCAAG	1843
misc_feature		Db	162397	CTCAGACAGCAGAGGTGGTTCCTGCCAGGAGCTCATGTATAGTGCAGGCGCTTGTGAGG	162338
misc_feature		Qy	1844	AGCCCATCGAAGGAGAGCTGTGGAGCCAGGACTCTGCCCAACAAAGACTTAGCTAAGCAGT	1903
misc_feature		Db	162337	GAAGCAGCCACCGCGCGCGGGGCTGTGTGCC-----	162305
misc_feature		Qy	1904	GACCAGAACCCACAAAACACTAGGCTGCATTTGTTGAAGAGGCAATCATTTTGGCAATTT	1963
misc_feature		Db	162304	-----CACTGAGAGGACAAAGTACGCTGGTGTCTCATCT	162272
misc_feature		Qy	1964	GTGAAAGTGTGTGATTTAGTAAATAATGTGAATAGCTTTGTACTATTATTTTGAAGACT	2023
misc_feature		Db	162271	GTGAA---TGTGTGGACTCCACAAATGTGAATATATTTTGTACAGATTTGGAAACT	162215

Query Match	31.8%;	Score 906;	DB 2;	Length 288118;
Best Local Similarity	69.3%;	Pred. No. 8.4e-138;		
Matches 1549;	Conservative 0;	Mismatches 560;	Indels 127;	Gaps 18;
QY	468	TATGATTATGATTAATTTGCCAGGACATCTGTTATCATAGCATTTTATAATGAAGCCTGG	527	
Db	163708	TATGATTATGATTAACCTTGCCCAAGACATCCGTTGTATAGCATTTTATAATGAAGCCTGG	163649	
QY	528	TCAACTCTCTTCGGACAGTTTACAGTGTCTCTTGAGACATCCCCGGATATCTCTGCTAGAA	587	
Db	163648	TCCACACTCTTTCAGACAGTTTACAGTGTCTCTGAGACTTCCCGTGATATCTGCTGGAG	163589	
QY	588	GAAAGTGATCTTGTAGATGACTACAGTGATAGAGACCTGAAGGAGCGCTTGGCCAAAT	647	
Db	163588	GAGGTCAATTTTGGTAGATGACTACAGTGACAGAGATCACCTGAAGGAGCGCCCTGGCCAA-	163530	
QY	648	GAGCTTTCCGGAGCTGCCCAAGTGCCTGATCGCGCCCAACAGAGAGAGGCGCTGGTG	707	
Db	163529	GAGCTGTACAGCTGCCCTTAGGTGCACCTGATCCCTGCTAGCA--AGAGAGGCGCTAGTG	163472	
QY	708	CGAGCCCGGCTGCTGGGGCGTCTGCGGCGAGGGGCGATGTTCTGACCTTCTCTGGACTGT	767	
Db	163471	CAAG-CCGGCTACTGGGAGCGTTTTCGCGCCAGGGGCGAAGTGTGAGGTTTCTGGACTGT	163413	
QY	768	CAGTGTGAGTGCCACGAAGGTTGCTGGAGCGGCTGCTGCAGAGGATCCATGAAGAGAG	827	
Db	163412	CAGTGTGACTGTCTATGAGGGTGTGCTGAGTCCCTGCTGCGAGAGGATCCACGAGAAGGAG	163353	
QY	828	TCGGCAGTGTGTGCCCGGTGATTGATGTGATCGACTGGAAACACTTCGAAATACCTGGGG	887	
Db	163352	TCGGCCGTGTGTGCCCGGTTATCGATGTCTATTGACTGGAAACACTCTGAGTACCTGGGC	163293	
QY	888	AACTCCGGGAGGCCCAAGATCGCGGTTTTCGACTGGAGGCTGGTGTTCACGTGGGCACACA	947	



```
QY 2024 TTTTAAATGTTCCAAATACCTATTTTCAAAGGGTAATCGTAAGATGTTAAACCTTGGT 2083
D 162214 TTTAAATATGCTTCCAGTATCCCGTTTCTAAGGCGGCTAGATGCTGCTTGGTA 162155
QY 2084 ATTTAGAAAATTAAACCTTTATATATTTTCTATCAARAWAWATTTTACAGTCGTGC 2143
D 162154 TTTGGGAAGCTTAATGCACGATGTCAT-----TTTACCCATGTCG 162113
QY 2144 CTTTACTCTCATTTAGCAAAAAGATAAAGATTTTATTTTGGTATTTTACAGAAATCCCA 2203
D 162112 CTTTATTTCTCATTTAGCAAAAAGGGAATATTTTACTTTTGTATTTACGAAGCTTCCCG 162053
QY 2204 GGTACGAAGATATCTGCATGGGTGGAATCAGGTTCAAGCAACGCTACTTTTGCACTTACTG 2263
D 162052 GGTGTTTATGATATCTTGA-----CACAACTGGTAAATACCTCAAAATGGGACTG 162004
QY 2264 ATATACCTCA--GCTGCGGGGTTAAAGTTTCCAGTATAGAGAGACTGTCACATAGGAA 2321
D 162003 GTATACCTCAATGTCATCAAGTTAGAGTTCTCTCAGGGGATGAGATGGACACTAGAAA 161944
QY 2322 CATTTGATTTGATTTTACG-----GTCAATGAGATCTTCTAGATGTTATTTTAAAGA 2375
D 161943 TACTGTCTTAATTTTGTGTTGTTGAGTGGGTGGGACAGTCTAGATTTTATCAAGAAGA 161884
QY 2376 ATGCTTTTGTGTTATGTTGTTGCTACACAGTTTACACTC---CATAATGTTTCAATGTCAGC 2432
D 161883 ATGCTTTTGTGTTATGTTGTTGCTATGCCACAGTCAAACTCCATCAATACGCTCAGACTGAC 161824
QY 2433 CAAAGAGACTTAACAAAGCTGAAATCTCAGAG--AACAAATTTGCTTTTACTAAGCTGAGTC 2491
D 161823 CCAAACAGGCTCACCATCTCTGAAAGCTCAAGAAAGATATGTTTGGCCAGGCTCTAC 161764
QY 2492 AACTT---GAGAGCGAATCTTCAATATGCC-----GCATGATGTTGTCGCTG 2536
D 161763 AAATCTGCGAGTACATTTTCCAGCAATGTCTATGGGAGGATGCTAGTCACTTTAGGAGG 161704
QY 2537 GTTCTACCATCATGACTTTTAAACATGTTTATATCATTTTAAATTTT-ATGATPACGTA 2595
D 161703 ACTTAACGACTGTGACTGGAAGAAAGATTCATATCACTCTAGTCTTTTAAATCAAGACA 161644
QY 2596 GTGTCAGGAGAAATGTAATGTTCTATATGAAATTCCTTTTCAAGTTTGTTCATTAATA 2655
D 161643 GTTCCAGGAGAAACGTAATGTTCTACATGAACTCACTTTTAAAGTTTTCCTTAATA 161584
QY 2656 ACAGTTATTAATTTAA 2671
D 161583 CTATTTGATTTGTTAA 161568

RESULT 14
AC099237_3
WPCOMMENT
Sequence split into 6 fragments LOCUS AC099237 Accession AC099237
Fragment Name Begin End
AC099237_0 1 110000
AC099237_1 100001 210000
AC099237_2 200001 310000
AC099237_3 300001 410000
AC099237_4 400001 510000
AC099237_5 500001 545260
Continuation (4 of 6) of AC099237 from base 300001 (AC099237 Rattus norvegicus clone CH2

Query Match 31.1%; Score 887.4; DB 2; Length 110000;
Best Local Similarity 69.6%; Pred. No. 8.1e-135;
Matches 1556; Conservative 0; Mismatches 551; Indels 128; Gaps 21;
QY 468 TATGATTTATGATTAATTTCCAGGACATCTGTTATCATAGCATTTTATTAATGAAGCCTGG 527
D 43593 TATGATTTATGATAACCTGCCAAGACATCCCGTTGTGTCATAGCATTTTATTAATGAAGCCTGG 43652
QY 528 TCAACTCTCTCTCGGACAGTTTACAGTGTCTCTGAGACATCCCGGATATCTCTGCTAGAA 587
```

```
D 43653 TCACACTCCTTTCAGACAGTTTACAGTGTTCGAGACTTCCCTCGATATCCCTGCTGGAG 43712
QY 588 GAAGTCATCTCTCTAGATGACTACAGTATAGAGACACCTGAAGAGCGCTTTGGCCAAT 647
D 43713 GAGGTCAATCTCTGTAGTACATGACTACAGTACAGAGATCACTGAAGAGGCCCTTGGCAA- 43771
QY 648 GAGCTTTTGGGACTGCCCCAAGGTGCGCTGATCCGCGCCCAACAAAGAGAGAGGCGCTGGT 707
D 43772 GAGCTGTACACAGCTGCCCCAGGTGACCTGATCCCTGGTAGCA--AGAGAGGCGCTAGTG 43829
QY 708 CGAGCCCGGCTGCTGGGGGGCTCTGGCGAGGGGGGATGTTCTGACCTTCTCTGGACTGT 767
D 43830 CAAG-CCGGCTACTGCGGAGCTCTGGGGCCAGGGGGAAGTGTGAGGTTTCTGGACTGT 43888
QY 768 CACTGTGAGTGCACCAAGGCTGCTGAGCGCTGCTGAGAGCGCTGCTGAGAGGAG 827
D 43889 CACTGTGACTGTCAATGAGGGGTGGCTGGAGT---CTGCTGAGAGGATCCACAGAGGAG 43946
QY 828 TCGGCAGTGTGTGCCCCGGTGTGATGTGATCGACTGGAACACCTTTCGAATACCTGGGG 887
D 43947 TCGGCCGTGGTGTGCCCGGTTATCGATGTCAATGACTGGAACACTTCTGAGTACCTGGGC 44006
QY 888 AACTCGGGAGAGCCCGAGATCGCGGTTTCGATCGAGCTGGAACACCTTTCGAATACCTGGGC 947
D 44007 AACTCC--GGAAGCCCGAGATTCGAGGCTTTGACTGGCGCTAGTATTTACGTGGCACGTG 44065
QY 948 GTTCTCGAGAGGAGAGGATACGATGCAATCCCGCTGATGTCATCAGGTCTCCAACA 1007
D 44066 GTCCCTTGGCGGAGCGGAAGTTGATGCGGACTCTTATGACATTAATCAGGTCTCCNACT 44125
QY 1008 ATGCTGTGGTGGCTGTTTGTGCTGTGAGTAAGAAATATTTTGAATATCTGGGGTCTTATGAT 1067
D 44126 ATGCTGTGGGACTGTTTGTGTTGAGTAAGAGATATTTTGAATACCTGGGGTCTTATGAT 44185
QY 1068 ACAGGAATGGAATGTTGGGAGAGAGAAACCTCGAATTTTCCCTTTAGGATCTGGCAGTGT 1127
D 44186 ACCGGAATGGAATCTAGGAGAGAGAAACCTTGAATTTCTCTTTAGGATCTGGCAATGT 44245
QY 1128 GGTGGGTTCTGGAACACACACACCATGTTCCCATGTTGGCCATGTTTCCCAAGCAAGCT 1187
D 44246 GGTGACTCTCTGGAACACACACCTTACTCCACGTGGGCCACGCTCTTCCCTAAGCAAGCT 44305
QY 1188 CCCTACTCTCCGCAACAGGCTCTGGCCAAACAGTGTTCGTGCACTGAAATGATGATGGAT 1247
D 44306 CCCTATTTACGCAAGCAAGGCTCTAGCCAAACAGTGTCCGAGCTGCAGAAAGTGGGCGAT 44365
QY 1248 GAATTTAAAGAGCTCTATACCATCGCAACCCCGCTGCGCGCTTGGAAACCTTTTGGGGAT 1307
D 44366 GAACTTTAAAGAACTCTACCAACCCCGGAAATCCCC--AGCCCGCTGGAAACCTCTTGGGGAG 44424
QY 1308 GTGACAGAGAGGAGCAGCTCCGGGCAACAGCTCCAGTGTAAAGACTTCAAGTGTCTTCTG 1367
D 44425 CTGACAGAGAGAAAGAAAGCTTTCAGCCCAAGCTCCAGTGTGAAGATTTAAAGCGGTTCTG 44484
QY 1368 GAGACTGTGTATCCAGAACTGCAATGCTGCTGAGGACAGGCTGCTTCTTCTCGGATGCTC 1427
D 44485 GCTTACTGTACCCAGNACTGCAATGCTGCAAGAGCAAGGCTGTTCTTCTCGGAAGCTT 44544
QY 1428 CAGAACAAAGGACTAAACAGACTACTGCTTTTGAATATAACCTTCCCGATGAAAACAGATT 1487
D 44545 GAGAACAGAGACTTACGGGGTACTGCTGGAATATAATCTCTCCAGTGAACAAATGTG 44604
QY 1488 GTGGGACACGAGTCAATCTGTACCTCTGTATGCTGGATGGCGGAGATCAGTTTTCGAG 1547
D 44605 GAAAGGCCACACGAGTCTTTTGTACCTCTGCGAGGGATGGTTCAGAACCAAGTTTTCGAG 44664
QY 1548 TACAGCTCCAGAGAAATAACGCTATAACCCACACAGCTCTGAGGCTCTGAGGCTGCTGCTGTG 1607
D 44665 TGTACATCCAGCAAGAAATACGGTATAACA--TCGCCAGCCCGCAGGCTCTGATAGCAGTG 44723
QY 1608 GAAGCAGGAATGGAATACCTTTATCATCTCTCTGCGAAGAACTGCCCGCAGAGAAATCAG 1667
D 44724 GAGGAAGGAAGGAT--CCTTATCCCGGCTCTCTGGAAGAGATACCGTTCAGAGAAATCAA 44781
```







QY 1128 GGTGGGTTCTGGAAACACACCCATGTTCCATGTTGGCCATGTTTCCCAAGCAGCT 1187  
DB 83964 GGTGACTCTCTGNAACACACCCCTTACTCCAGTGGCCACAGCTCTTCCCTTAAGCAGCT 83905  
QY 1188 CCTACTCCCGCAACAAAGGCTCTGGCAACAGTGTTCGTGCAGCTGAAGTATGGATGGAT 1247  
DB 83904 CCTATTACGACGACGAGGCTCTAGCCCAACAGTGTCCGAGCTGCAGAAGTGTGGCGGAT 83845  
QY 1248 GAATTTAAAGAGCTCTACTACCACTCGCAACCCCGTCCCGCTTGGAAACCTTTTGGGGAT 1307  
DB 83844 GAACTTTAAAGAGCTCTACTACCAACCCCGGAAATCCCC - AGCCGCGCTGGAAACCTCTCGGGAG 83786  
QY 1308 GTGACAGAGAGGAGCAGCTCCGGGCAAGCTCCAGGTGAAGACTTTCRAAGTGTCTTCG 1367  
DB 83785 CTGACAGAGAGAGAGAGCTTTCAGCCCAAGCTTCAGTGTGAAGATTTAAAGCGGTTCCG 83726  
QY 1368 GAGACTGTGTATCCAGAACTGTCATGTGCTGAGGACAGGCTGCTCTTTCGGGATGCTC 1427  
DB 83725 GCTACACTGTACCCAGAACTGTCATGTGCCAGAGACAGGCTGCTTCTTCGGGAGCTT 83666  
QY 1428 CAGAACAAAGGACTAACAGACTACTGCTTTGACTATTAACCTCCCGATGAAGAACAGAT 1487  
DB 83665 GAGAACAGAGAGGACTACGGGGGTACTGCTCGGACTATAATCTCCAGTGAAGAACTATGTG 83606  
QY 1488 GTGGGACACCAAGTCTTCTGTACCTCTGTCATGGGATGGCCAGATCAGTTTTCGAG 1547  
DB 83605 GAAGGCCACCAAGTCTTCTGTACCTCTGTCAGCGGATGGGTGAGAACCAAGCTTTTCGAG 83546  
QY 1548 TACACGTCCTCCAGAAAGAAATACGCTATAACCCACCCAGGCTCAGGGCTCATTGTGCTG 1607  
DB 83545 TGTACATCTCCAGCAGAAATACGGTATAACA - TCGCCAGCCCGAGGCTGTCATAGCAGTG 83487  
QY 1608 GAAGCAGGAATGATACCTTTATCATGTCATCTCTGCGAAGAACTCTGCCCGCAGAGATCAG 1667  
DB 83486 GAGGAAGGGAAGGAT - CCTTATCCCGGCTCTCTGGAAGATACCGTTCCAGAGAACTCA 83429  
QY 1668 AGGTTTCATCTTGAGGAGGATGATCTTTATTTTCAGAACAGTCCAGAAATGTGCCAG 1727  
DB 83428 GAGTTTCATCTCAGGAGGACCGACCTTAGTCTCAAGCAGAGCAGGAAATGTGGAG 83369  
QY 1728 GCTGCGAGGAAGGAGTGCAGTGCAGTTCGTTCCACTTTACGAGACTGCACCAACTCG 1787  
DB 83368 GCCACAGAGGAGGATTTCAACTCTGCTTGGCTCCGAACTTGGAGACTGTAGCAACTCA 83309  
QY 1788 G - ATCATCAGAAATGGTCTTCAAGAGCGCATGTTATGAAGCCTCGTGTATCAAGGAG 1845  
DB 83308 GACACAGCCAGAGTGTGCTTCAAGGAGTGATGTATAGTGCACGGCTTGTGAGG - 83251  
QY 1846 CCATCGAAGGAGACTGTGGAGCCAGACTCTGCCCAACAAAGACTTAGCTAAGCAGTGA 1905  
DB 83250 - - - - - GAAGGAGCCCAACCGCGCCGGGGCTG - - - - - 83224  
QY 1906 CCAGAACCCCAAAACTAGGCTGATGCTTTTGAAGGGCAATCATTTTGGCAATTTGT 1965  
DB 83223 - - - - - TGTCCCAAGTGAGAGGCAAGTACGTTGTCATCTGT 83183  
QY 1966 GAAAGTGTGTGGATTTAGTAAATAATGGAATTTGTAATTTTGTAGAACTTT 2025  
DB 83182 GAA - - - TGTGTGGACTCCACAACTATGAAATTTATTTGTACAGATTTGGAAACTTT 83126  
QY 2026 TTAATGTTCAAAATACCTATTTCGAAAGGTAATCGTAAGATGTTAACCCCTGGGTAT 2085  
DB 83125 AAAAATGTCTTCCAGTATCCGTTTCTAAGGGCCGCTCTAAGATGCTGCTTTGGTATT 83066  
QY 2086 TTAGAAAAATTAACCTTATAATTTTCTATCAARAWAWATTTTACAGTCTGCTT 2145  
DB 83065 TGGGAAGCTTAATGCAGATGTCAT - - - - - TTACCATGGTGCCT 83024  
QY 2146 TTTACTCTCATTAGCAAAAAAGATAAAGATTTTATTTTGGTATTTTACAAGAAATTTCCAGG 2205  
DB 83023 TTTATTTCTCATTAGCAAGAGGGAATATTTTACTTTGTACTTACGAGCTTCCGGG 82964  
QY 2206 TAGGAAGATATCTGCATGGGTGGAAATCAGGTTTCAAGCAACGACTTTTGCATTAACATGAT 2265

DB 82963 TGTTTAGTATACTTGA - - - - - CAAAACTGGTAATACTCTCAAAATGTGGACTGGT 82915  
QY 2266 AATACCTCA - - GCTGGCGGGTTAAAGTTTCCAGTATAGAGAGACTGTCACTAGGAACA 2323  
DB 82914 AATACCTCAATGTCATCAAGTTAGAGTCTCTCAGGGGATGAGTGCACACTAGAATA 82855  
QY 2324 TTGTATTGATTTATTCAG - - - - - GTCATTTGAGATCTCTTAGATGTTATTTTAAAGAAAT 2377  
DB 82854 CTGCTTAATTTTGGTTGTGAGTGGGTGGACAGTCTAGATTTATACAAGAAAGAAAT 82795  
QY 2378 GCTTTTGGTTATGCTGTGCTACCAAGTAAACACTC - - - - - CATAAATGTTTCATGTCAGCCA 2434  
DB 82794 GCTTTTGTCTATGCTGATGCGCAGTCAAGTCAAACTCCATCATACGCTCAGACTGACCC 82735  
QY 2435 AAGAGGACTAAACCAAGCTGAAATCTCAGAG - AACAAATTTGCTTTTACTAAAGCTGAGTCAA 2493  
DB 82734 AACAGGCTCACCATCTCTGAAAGCTCAAGAAAGAAATATGGTTTGGCAAGGTCCTACAA 82675  
QY 2494 CTT - - - - - GAGAGCGAACTTCTAACAATGCCGACTGTAGTGTGCTGCT - - - - - 2538  
DB 82674 ACTTGGAGTGTACATTTCCAGCAATGCTAGTGGGAGGATGCTAGTACTTACGAGGA 82615  
QY 2539 -TCTACCACTATGACTTTAAACATGTTTATATCATTTTAAATTTT - ATGATACGGTAG 2596  
DB 82614 CTTAACGACTGTGACTGNAAAAAAGCAATTCAGATCATTTCTAGTTTAAATCAAAGGACAG 82555  
QY 2597 TGTGCGGAGAAATGTAATGTTCTATATGAAATTCCTTTTCAAGTTTGTTCATTAATAA 2656  
DB 82554 TTCCAGGAGAAACGTAATGTTCTACATGAAGTTCGACTTAAAGTTTTTCTTAATAA 82495  
QY 2657 CAGTTAATTAATTA 2671  
DB 82494 TATTTGATTTTAA 82480

Search completed: December 7, 2003, 18:37:45  
Job time : 10426.6 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 7, 2003, 11:24:22 ; Search time 736.064 Seconds  
(without alignments)  
10452.079 Million cell updates/sec

Title: US-10-074-527-1  
Perfect score: 2850  
Sequence: 1 cggctcggtaccactataa.....aaaaaaaaaaaaaaaaaaaaa 2850

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Genes19Jun03.\*

```
1: /SIDSL1/gcgdata/geneseq/geneeqn-emb1/NA1980.DAT.*
2: /SIDSL1/gcgdata/geneeq/geneeqn-emb1/NA1981.DAT.*
3: /SIDSL1/gcgdata/geneeq/geneeqn-emb1/NA1982.DAT.*
4: /SIDSL1/gcgdata/geneeq/geneeqn-emb1/NA1983.DAT.*
5: /SIDSL1/gcgdata/geneeq/geneeqn-emb1/NA1984.DAT.*
6: /SIDSL1/gcgdata/geneeq/geneeqn-emb1/NA1985.DAT.*
7: /SIDSL1/gcgdata/geneeq/geneeqn-emb1/NA1986.DAT.*
8: /SIDSL1/gcgdata/geneeq/geneeqn-emb1/NA1987.DAT.*
9: /SIDSL1/gcgdata/geneeq/geneeqn-emb1/NA1988.DAT.*
10: /SIDSL1/gcgdata/geneeq/geneeqn-emb1/NA1989.DAT.*
11: /SIDSL1/gcgdata/geneeq/geneeqn-emb1/NA1990.DAT.*
12: /SIDSL1/gcgdata/geneeq/geneeqn-emb1/NA1991.DAT.*
13: /SIDSL1/gcgdata/geneeq/geneeqn-emb1/NA1992.DAT.*
14: /SIDSL1/gcgdata/geneeq/geneeqn-emb1/NA1993.DAT.*
15: /SIDSL1/gcgdata/geneeq/geneeqn-emb1/NA1994.DAT.*
16: /SIDSL1/gcgdata/geneeq/geneeqn-emb1/NA1995.DAT.*
17: /SIDSL1/gcgdata/geneeq/geneeqn-emb1/NA1996.DAT.*
18: /SIDSL1/gcgdata/geneeq/geneeqn-emb1/NA1997.DAT.*
19: /SIDSL1/gcgdata/geneeq/geneeqn-emb1/NA1998.DAT.*
20: /SIDSL1/gcgdata/geneeq/geneeqn-emb1/NA1999.DAT.*
21: /SIDSL1/gcgdata/geneeq/geneeqn-emb1/NA2000.DAT.*
22: /SIDSL1/gcgdata/geneeq/geneeqn-emb1/NA2001A.DAT.*
23: /SIDSL1/gcgdata/geneeq/geneeqn-emb1/NA2001B.DAT.*
24: /SIDSL1/gcgdata/geneeq/geneeqn-emb1/NA2002.DAT.*
25: /SIDSL1/gcgdata/geneeq/geneeqn-emb1/NA2003.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2848	99.9	2850	24	ABSS2469 Human cDNA encodin
2	2738.2	96.1	2745	24	AAD40568 Human drug metabol
C 3	2081.8	73.0	2290	22	AAI26522 Human breast cancer
C 4	2081.8	73.0	2290	22	AAI26538 Human breast cancer
5	1208.4	42.4	1259	21	AAC98209 Human colon cancer
C 6	1115.2	39.1	1352	23	ABV23465 Human prostate exp
C 7	1115.2	39.1	1352	23	ABV29323 Human prostate exp
8	1012.2	35.5	1517	22	AAH33532 Human colon cancer

*same as in gene emb1*

9	1012.2	35.5	1517	24	ABL90290 Human polynucleoti
10	1011.8	35.5	1517	21	AAA78414 DNA encoding novel
11	967.4	33.9	1069	23	AAS68954 Human breast cancer
C 12	552.2	26.5	883	22	AAI26610 Human colon cancer
13	552.8	19.4	592	21	AAA02624 Human ovarian tumo
14	527	18.5	554	22	AH82814 Human ovarian PCR-
15	526.4	18.5	544	22	AAS24233 Human ovarian tumo
16	497	17.4	652	22	AH82556 Human ovarian PCR-
17	496.4	17.4	632	22	AAS23991 Human ovarian tumo
18	496.4	17.4	632	22	AAS25242 Human N-acetylglala
19	496.2	17.4	1737	25	ABX14962 Human N-acetylglala
20	450.8	15.8	496	24	ABL81812 Human breast cancer
C 21	445.4	15.6	470	22	AAI16882 Human breast cancer
C 22	443.4	15.6	473	24	ABU81811 Human ovarian can
C 23	425.8	14.9	473	24	AAU80705 Human ovarian can
C 24	425	14.9	437	22	AAI25525 Human prostate exp
25	423	14.8	425	23	ABV14650 Human prostate exp
26	423	14.8	471	23	ABV35728 Human prostate exp
27	423	14.8	471	23	ABV4532 Human prostate exp
28	422	14.8	439	23	ABV05481 Human breast cancer
C 29	419.2	14.7	450	22	AAI16580 Human breast cancer
C 30	415.4	14.6	417	22	AAI25423 Human breast cancer
C 31	409	14.4	421	24	ABL62213 Colon adenocarcino
C 32	409	14.4	421	24	ABL67829 Ovary cancer relat
C 33	406.4	14.3	408	24	ABL87421 Human ovarian can
C 34	363	12.7	483	24	ABK53538 Human eosinophil-m
C 35	361	12.7	414	22	AAI07779 Human breast cancer
C 36	343.8	12.1	384	22	AAI07677 Human breast cancer
37	285.6	10.0	1920	22	AAF63512 UDP-GalNac: polype
38	285.6	10.0	2102	24	ABQ99468 Human coding seque
39	285.6	10.0	2575	21	AAA37122 Human PRO1564 (UNQ
40	285.6	10.0	2575	22	AA46106 Human DNA encoding
41	285.6	10.0	2575	22	AAAF4445 Primer #98 used in
42	285.6	10.0	2575	25	ACA57864 Human PRO1564 CDNA
43	285.6	10.0	2575	25	ABX98334 Human cDNA encodin
44	285.6	10.0	2575	25	ABX98836 Novel human secret
45	285.6	10.0	2575	25	ACA05881 Human secreted/tra

#### ALIGNMENTS

RESULT 1  
ABSS2469  
ID ABSS2469 standard; cDNA; 2850 BP.  
XX AC ABSS2469;  
XX AC  
XX DT 15-NOV-2002 (first entry)  
XX DE Human cDNA encoding glycosyltransferase 33945.  
XX DE Human; ss; gene; glycosyltransferase; 33945; atherosclerosis;  
XX KW cardiovascular disorder; ischaemia; atherosclerosis; cancer; tumour;  
XX KW congestive heart failure; endothelial cell disorder; psoriasis;  
XX KW diabetic retinopathy; angina; hypertension; atrial fibrillation;  
XX KW valvular disease; cardiomyopathy; haemangioma; pancreatic disorder;  
XX KW cellular proliferative disorder; differentiative disorder; diabetes;  
XX KW autoimmune disorder; haematopoietic neoplastic disorder; leukaemia;  
XX KW Hodgkin's disease; chronic myelogenous leukaemia; inflammatory disease;  
XX KW arthritis; multiple sclerosis; viral infection; liver disorder;  
XX KW liver fibrosis; hepatocellular cancer.

OS Homo sapiens.

PH Key Location/Qualifiers  
FT CDS 81..1826  
FT /tag= a  
FT /product= "Glycosyltransferase 33945"  
FT /note= "this CDS is specifically claimed in claim 1"

XX WO200264815-A2.





(INCY-) INCYTE GENOMICS INC.

Sanjanwala MM, Yao MG, Au-young J, Baughn MR, Arvizu C, Ring HZ;  
Lee EA, Ding L, Hafalia AJA, Tang YT, Yue H, Triboulet CM;  
Lu DAM, Lal PG, Warren BA, Yang J, Walia NK, Nguyen DB;  
Gandhi AR, Lu Y, Ison CH;

WPI; 2002-519668/55.  
P-PSDB; AAE25019.

Novel human drug metabolizing polypeptide, useful in diagnosis, prevention or treatment of autoimmune/inflammatory, cell proliferative, neurological, developmental, endocrine, metabolic and gastrointestinal disorders -

Claim 72: page 161-162; 169pp: English.

The invention relates to an isolated human drug metabolising enzyme (DME) and its nucleotide. DME is useful for diagnosing, treating or preventing disorders associated with aberrant expression of DME, where the disorders are selected from autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), asthma, atherosclerosis, psoriasis, uveitis; a cell proliferative disorder such as arteriosclerosis, cirrhosis, hepatitis, and cancer; a neurological disorder such as Alzheimer's disease, Huntington's disease, dementia, Parkinson's disease; a developmental disorder such as renal tubular acidosis, epilepsy, anaemia; an endocrine disorder such as adenoma, thrombosis and infections; an eye disorder such as conjunctivitis, glaucoma, cataract; metabolic disorder such as cystic fibrosis, diabetes and goitre; a gastrointestinal disorder such as anorexia, peptic ulcer; and liver disorders. DME is useful in a number of drug screening techniques and to analyse the proteome of a tissue or cell type. The invention is useful for creating knock-in humanised animals or transgenic animals to model human diseases, in somatic or germline gene therapy, to generate a transcript image of a tissue or cell type, for detecting differences in the chromosomal location due to translocation, inversion, etc. among normal, carrier or affected individuals, and as hybridisation probes for mapping naturally occurring genomic sequences. The present sequence is human DME-4 cDNA.

Sequence 2745 BP: 704 A: 612 C: 737 G: 692 T: 0 other:

Query Match 96.1%; Score 2738.2; DB 24; Length 2745;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 2737; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

81 ATGTGGGGCGCAGCGCGGGCGGCTGCCCGGGGAATGCGGCGCGCGGAGGCG 140

db  
1 ATGTGGGGCGCAGGGCGGGCGCGCTGCCGGGAACTGCCGGCGGCCCGGAGGCG 60

141 CTGTTGGTGCTCCTGGCGGCTA CTGGCGTGGCCGGCTCGGTGCTGCGGGCCGAG 200

61 CTGTTGGTGGTCTCTGGGGTAA CTGGCGTGGCGGCTGGGTGCTCGGGCGAG 120

DV 201 CGTGGGCGCGGGCGGGGCTGCCGAGCGCGGAACCCGCGGCACCGCGGGCGG 260

db 121 CGTGGGGCCGGGGCCGGGCTGCGGAGCCCGGACCCCGGCGCCGGGCGG 180

261 CCGAGCCGGTCA TCCCGCGCGCCGCGTGGCGGGA CGCGCTGGCGCGGGCGAG 320

db : 181 CCGAGCCCGCTCATGCCGCGGGCCGCCCGCTCCGGCGAACCGCGTGGCGGCGCGGGCGAG 240

[illegible]

nb 241 300

2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525  
 526

2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525  
 526

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525  
 526



```
QY 1581 CACCAGCTGAGGCTGCATTGCTGTGGAAGCAGGAATGGATACCTTATCATGCAATCTC 1640
Db 1501 CACCAGCTGAGGCTGCATTGCTGTGGAAGCAGGAATGGATACCTTATCATGCAATCTC 1560
QY 1641 TGCAGAGAACTGCCCCCAGAGAAATCAGAAATTCATCTTGAGGAGGATGGATCTTTATTT 1700
Db 1561 TGCAGAGAACTGCCCCCAGAGAAATCAGAAATTCATCTTGAGGAGGATGGATCTTTATTT 1620
QY 1701 CACGAAAGTCCAGAAATGTTCCAGGCTGCGAGGAGGAGTGCAGTGACAGTTTCGTT 1760
Db 1621 CACGAAAGTCCAGAAATGTTCCAGGCTGCGAGGAGGAGTGCAGTGACAGTTTCGTT 1680
QY 1761 CCATCTTTACGAGACTGCACCACTCGGATCATCAGAAATGGTTCTTCAAGAGCGCATG 1820
Db 1681 CCATCTTTACGAGACTGCACCACTCGGATCATCAGAAATGGTTCTTCAAGAGCGCATG 1740
QY 1821 TTATGAAGCCTCGTGTATCAAGGAGCCCATCGAAGGAGACTGTGAGCCAGGACTCTGCC 1880
Db 1741 TTATGAAGCCTCGTGTATCAAGGAGCCCATCGAAGGAGACTGTGAGCCAGGACTCTGCC 1800
QY 1881 CAACAAAGACTTAGCTAAGCAGTGCAGAGAACCCACCAAACTAGGCTGCAATTTGTTG 1940
Db 1801 CAACAAAGACTTAGCTAAGCAGTGCAGAGAACCCACCAAACTAGGCTGCAATTTGTTG 1860
QY 1941 AAGAGGCAATCATTTTCCCATTTGTAAGTTGTTGGATTTAGTAAATGTAATAA 2000
Db 1861 AAGAGGCAATCATTTTCCCATTTGTAAGTTGTTGGATTTAGTAAATGTAATAA 1920
QY 2001 GCTTTGACTTATTTTTCAGAACTTTTAAATGTTTCCAAATACCCCTATTTTCAAGGGTA 2060
Db 1921 GCTTTGACTTATTTTTCAGAACTTTTAAATGTTTCCAAATACCCCTATTTTCAAGGGTA 1980
QY 2061 ATCTAAGATGTTTAACTTGGTATTTAGAAATTTAAACCTTATAATATTTTCTATCA 2120
Db 1981 ATCTAAGATGTTTAACTTGGTATTTAGAAATTTAAACCTTATAATATTTTCTATCA 2040
QY 2121 ABAWRWAWATTTTACAGTCTGCTTTTACTCTCATTTAGCAAAAAGATAAAGATTTAT 2180
Db 2041 AGATGTATATTTTACAGTCTGCTTTTACTCTCATTTAGCAAAAAGATAAAGATTTAT 2100
QY 2181 TTTGGTATTTTACAGAAATTTCCAGGTACGAAGATATCTGATGGTGGAAATCAGGTTCA 2240
Db 2101 TTTGGTATTTTACAGAAATTTCCAGGTACGAAGATATCTGATGGTGGAAATCAGGTTCA 2160
QY 2241 AGCAACGTACTTTGCATTAATACCTGATACCTCAGCTGCGGGTTAAAGTTTCCCAAGTA 2300
Db 2161 AGCAACGTACTTTGCATTAATACCTGATACCTCAGCTGCGGGTTAAAGTTTCCCAAGTA 2220
QY 2301 TAGAGAGACTGTCACTAGGAACATTTGATTTATTTTACAGTCAATTCAGATCTCTAGA 2360
Db 2221 TAGAGAGACTGTCACTAGGAACATTTGATTTATTTTACAGTCAATTCAGATCTCTAGA 2280
QY 2361 TGTATTTTAAAGAAATGCTTTTGGTTATGTTGCTTACACAGTTTAACTCCATAAT 2420
Db 2281 TGTATTTTAAAGAAATGCTTTTGGTTATGTTGCTTACACAGTTTAACTCCATAAT 2340
QY 2421 GTTCATGTCAGCCAAAGAGACTAACCAAAAGCTGAAATCTCAGAGAAATTTGCTTTAC 2480
Db 2341 GTTCATGTCAGCCAAAGAGACTAACCAAAAGCTGAAATCTCAGAGAAATTTGCTTTAC 2400
QY 2481 TAAGCTGAGTCACTTGAAGCGAACTTTCTAAATCCCGCACTGTAGTGGCTGTTTC 2540
Db 2401 TAAGCTGAGTCACTTGAAGCGAACTTTCTAAATCCCGCACTGTAGTGGCTGTTTC 2460
QY 2541 TACCACATGACTTTAAACATGTTTATATCATTTTAAATTTTATGATACGGTAGTGTTC 2600
Db 2461 TACCACATGACTTTAAACATGTTTATATCATTTTAAATTTTATGATACGGTAGTGTTC 2520
QY 2601 AGGAGAAATGTAATGTTCTATATGAATTTCTTTTCAAGTTTGTTCATTAATAACAGT 2660
Db 2521 AGGAGAAATGTAATGTTCTATATGAATTTCTTTTCAAGTTTGTTCATTAATAACAGT 2580
QY 2661 TATTAATTTAAATCAGCGTTAGAGTTTGTGCTGCTGCAACTGCTGTGAAAAATTTCTCTGA 2720
```

```
Db 2581 TATTAATTTAAATCAGCGTTAGAGTTTGTCTGCTGCAACTGCTGTGAAAAATTTCTCTGA 2640
QY 2721 GTAATTTCTGATTTGTGAATGATCCAGACCAACCTCGAGATTTTGTCAACTGATTAAGT 2780
Db 2641 GTAATTTCTGATTTGTGAATGATCCAGACCAACCTCGAGATTTTGTCAACTGATTAAGT 2700
QY 2781 CAATATGAATGATTAATAAAGATGTGAGAACAAAAA 2825
Db 2701 CAATATGAATGATTAATAAAGATGTGAGAACAAAAA 2745

RESULT 3
AAL26522/c
ID AAL26522 standard; cDNA; 2290 BP.
XX
AC AAL26522;
XX
DT 07-DEC-2001 (first entry)
XX
DE Human breast cancer expressed polynucleotide 18979.
XX
KW Human; breast cancer; cell marker; cytostatic; ss.
XX
OS Homo sapiens.
XX
PN WO200151628-A2.
XX
PD 19-JUL-2001.
XX
PF 10-JAN-2001; 2001WO-US000798.
XX
PR 14-JAN-2000; 2000US-0176077.
PR 14-MAR-2000; 2000US-0189167.
PR 24-MAR-2000; 2000US-0192099.
PR 29-MAR-2000; 2000US-0193480.
PR 15-MAY-2000; 2000US-0205230.
PR 09-JUN-2000; 2000US-0211315.
PR 25-JUL-2000; 2000US-0220534.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX
WPI; 2001-451856/48.
XX
PT New peptide useful as a marker for the diagnosis of breast cancer -
XX
PS Claim 1; Page 3534-3535; 3695pp; English.
XX
CC The invention relates to human breast cancer expressed polynucleotides
CC (AAL07544-AAL26789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterising treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
CC activity.
XX
SQ Sequence 2290 BP; 508 A; 676 C; 572 G; 528 T; 6 other;
XX
Query Match 73.0%; Score 2081.8; DB 22; Length 2290;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 2105; Conservative 1; Mismatches 3; Indels 2; Gaps 2;
QY 12 CCACTATAACGGCGCCGAGTGTGCTGGAATTCGCCCTTGGCAGATCGCTGGCTGCAGTT 71
Db 2280 CCACTATAACGGCGCCGAGTGTGCTGGAATTCGCCCTTGGCAGATCGCTGGCTGCAGTT 2222
QY 72 GCGCGGCGCATGTGGGGGCGCACGGCGCGCGCGCTGCCCGGGAACTGCGCGCGCGC 131
Db 2221 GCGCGGCGCATGTGGTGGGCGCACGGCGCGCGCGCTGCCCGGGAACTGCGCGCGCGC 2162
```

QY 132 CGGAGCGCTGTTGGTCTCTGGCGCTACTGCGCTTGGCCGGGCTGGGCTGGTCTG 191  
DB 2161 CGGAGCGCTGTTGGTCTCTGGCGCTACTGCGCTTGGCCGGGCTGGGCTGGTCTG 2102  
QY 192 CGGCGCAGCGTGGGCGCGGGCGCGGGCTGCGGAGCCGGGACCCCGCGCACCCCGCGC 251  
DB 2101 CGGCGCAGCGTGGGCGCGGGCGCGGGCTGCGGAGCCGGGACCCCGCGCACCCCGCGC 2042  
QY 252 CCGGGCGCGCAGCGCGTCTATGCCCGCGCGCGCGTGGCGGAAACGGCGCTGGGCGG 311  
DB 2041 CCGGGCGCGCAGCGCGTCTATGCCCGCGCGCGCGTGGCGGAAACGGCGCTGGGCGG 1982  
QY 312 CGGGCGAGCGGTCGGCTGCGAGTCCAGGGCGAGAGCTGCGGCTGCGAGGAGAGC 371  
DB 1981 CGGGCGAGCGGTCGGCTGCGAGTCCAGGGCGAGAGCTGCGGCTGCGAGGAGAGC 1922  
QY 372 GTGGCGCTGCACAGATTAACATCTACCTCAGCGACCGCATCTCACCTGCACCGCGCTG 431  
DB 1921 GTGGCGCTGCACAGATTAACATCTACCTCAGCGACCGCATCTCACCTGCACCGCGCTG 1862  
QY 432 CCGAGCGCTGGAACCCCGCTGTGCAAGAGAGAAATATGATTATGATTAATTTGCCGAG 491  
DB 1861 CCGAGCGCTGGAACCCCGCTGTGCAAGAGAGAAATATGATTATGATTAATTTGCCGAG 1802  
QY 492 ACATCTGTTATCATAGCATTTTATAATGAAGCCCTGCTCACTCTCTCGGACAGTTTAC 551  
DB 1801 ACATCTGTTATCATAGCATTTTATAATGAAGCCCTGCTCACTCTCTCGGACAGTTTAC 1742  
QY 552 AGTGTCTTGAACATCCCGGATATCTCTGTAAGAGAGTGTCTTGTAGATGACTAC 611  
DB 1741 AGTGTCTTGAACATCCCGGATATCTCTGTAAGAGAGTGTCTTGTAGATGACTAC 1682  
QY 612 AGTGTATGAGAGCACTGGAAGAGCGCTTGGCCAAATGAGCTTTCGGGACTGCCCAAGTG 671  
DB 1681 AGTGTATGAGAGCACTGGAAGAGCGCTTGGCCAAATGAGCTTTCGGGACTGCCCAAGTG 1622  
QY 672 CGCTGTATCCGCGCAACAGAGAGCGCTGCTGCGAGCCCGCTGCTGGGCGCTCT 731  
DB 1621 CGCTGTATCCGCGCAACAGAGAGCGCTGCTGCGAGCCCGCTGCTGGGCGCTCT 1562  
QY 732 CGCGCGAGGGCGATGTTCTGACCTTCTGACCTGTCTCACTGTGAGTGCACAGAGGCTG 791  
DB 1561 CGCGCGAGGGCGATGTTCTGACCTTCTGACCTGTCTCACTGTGAGTGCACAGAGGCTG 1502  
QY 792 CTGAGCGCGCTGTGCGAGAGATCCATGAAGAGAGTCCGAGTGTGTCGCCGCTGATT 851  
DB 1501 CTGAGCGCGCTGTGCGAGAGATCCATGAAGAGAGTCCGAGTGTGTCGCCGCTGATT 1442  
QY 852 GATGTATCGACTGGAACACCTTCAATACCTGGGAACTCCGGGAGCCCGAGATCGG 911  
DB 1441 GATGTATCGACTGGAACACCTTCAATACCTGGGAACTCCGGGAGCCCGAGATCGG 1382  
QY 912 GGTTCGACTGGAGCTGCTGTCACTGCGGACACAGTTCCTGAGAGGAGAGATACGG 971  
DB 1381 GGTTCGACTGGAGCTGCTGTCACTGCGGACACAGTTCCTGAGAGGAGAGATACGG 1322  
QY 972 ATGCAATCCCGCTCGATGTCATCAGGTCCTCAACAAATGCTGTGGGCTGTTGCTGTG 1031  
DB 1321 ATGCAATCCCGCTCGATGTCATCAGGTCCTCAACAAATGCTGTGGGCTGTTGCTGTG 1262  
QY 1032 AGTAAGAAATATTTGAAATATCTGGGCTTATGATACAGAAATGGAAGTTGGGAGGA 1091  
DB 1261 AGTAAGAAATATTTGAAATATCTGGGCTTATGATACAGAAATGGAAGTTGGGAGGA 1202  
QY 1092 GAAACCTCGAATTTTCTTTAGGATCTGGCAGTGTGGTGGGCTTCTGGAAACACACCA 1151  
DB 1201 GAAACCTCGAATTTTCTTTAGGATCTGGCAGTGTGGTGGGCTTCTGGAAACACACCA 1142  
QY 1152 TGTTCCTCATGTTGGCAGTGTGTTTCCCAAGCAAGCTCCCTACTCCCGCAACAGGCTCTG 1211  
DB 1141 TGTTCCTCATGTTGGCAGTGTGTTTCCCAAGCAAGCTCCCTACTCCCGCAACAGGCTCTG 1082

QY 1212 GCCAACAGTGTTCGTGCGAGCTGAAGTATGATGATGAATTTAAAGAGCTCTACTACCAT 1271  
DB 1081 GCCAACAGTGTTCGTGCGAGCTGAAGTATGATGATGAATTTAAAGAGCTCTACTACCAT 1023  
QY 1272 CGCAACCCCGCTGCGCGCTTGGAACTTTTGGGGATGTGACAGAGAGGAGAGCTCCGG 1331  
DB 1022 CGCAACCCCGCTGCGCGCTTGGAACTTTTGGGGATGTGACAGAGAGGAGAGCTCCGG 963  
QY 1332 GACAAGCTCCAGTGAAGAGCTTCAAGTGTCTTCTGGAGACTGTGTATCCAGAACTGCAT 1391  
DB 962 GACAAGCTCCAGTGAAGAGCTTCAAGTGTCTTCTGGAGACTGTGTATCCAGAACTGCAT 903  
QY 1392 GTCCCTGAGGACAGAGGCTGCGCTTCTTCTGGGATGCTCCAGAAACAAAGGACTTAAACAGACTAC 1451  
DB 902 GTCCCTGAGGACAGAGGCTGCGCTTCTTCTGGGATGCTCCAGAAACAAAGGACTTAAACAGACTAC 843  
QY 1452 TGTCTTGAATTAACCTTCCGATGAAACCCAGATTTGTGGGACACAGGTCATTCTGTATC 1511  
DB 842 TGTCTTGAATTAACCTTCCGATGAAACCCAGATTTGTGGGACACAGGTCATTCTGTATC 783  
QY 1512 CTCTGTATGGGATGGGCGCAGATCAGTTTCTCGAGTACACGTCCTCCAGAAAGAAATACGC 1571  
DB 782 CTCTGTATGGGATGGGCGCAGATCAGTTTCTCGAGTACACGTCCTCCAGAAAGAAATACGC 723  
QY 1572 TATAACACCCAGGCTGAGGCTGCTGCTGTGGAAGCAGAGAAATGGAATACCTTATC 1631  
DB 722 TATAACACCCAGGCTGAGGCTGCTGCTGTGGAAGCAGAGAAATGGAATACCTTATC 663  
QY 1632 ATGCATCTCTGCGAAGAACTGCCCCAGAGAAATCAGAAATTCATCTTTCAGAGGAGTGA 1691  
DB 663 ATGCATCTCTGCGAAGAACTGCCCCAGAGAAATCAGAAATTCATCTTTCAGAGGAGTGA 603  
QY 1692 TCTTTATTTTACGAAACAGTCCAGAAATGTCTCAGGCTGCGAGGAGGAGTGCAGTGAC 1751  
DB 602 TCTTTATTTTACGAAACAGTCCAGAAATGTCTCAGGCTGCGAGGAGGAGTGCAGTGAC 543  
QY 1752 AGTTTCGTTTCCACTCTTACGAGACTGCACCACTCGGATCATCAGAAATGTTCTTCAAA 1811  
DB 542 AGTTTCGTTTCCACTCTTACGAGACTGCACCACTCGGATCATCAGAAATGTTCTTCAAA 483  
QY 1812 GAGCGATGTTATGAAGCTGCTGTATCAAGAGGCCATCGAAGGAGACTGTGGAGCCAG 1871  
DB 482 GAGCGATGTTATGAAGCTGCTGTATCAAGAGGCCATCGAAGGAGACTGTGGAGCCAG 423  
QY 1872 GACTCTGCCCAACAAAGACTTACCTAAGCAGTGCACAGAACCCACCAAACTAGGCTGC 1931  
DB 422 GACTCTGCCCAACAAAGACTTACCTAAGCAGTGCACAGAACCCACCAAACTAGGCTGC 363  
QY 1932 ATTGCTTTGAAGAGGCAATCATTTTGCATTTGTGAAAGTGTGTGGATTTAGTAAAAA 1991  
DB 362 ATTGCTTTGAAGAGGCAATCATTTTGCATTTGTGAAAGTGTGTGGATTTAGTAAAAA 303  
QY 1992 TGTGAATAAGCTTTGACTTATTTTGGAACTTTTGGAACTTTTAAATGTTTCCAAATACCTATTTT 2051  
DB 302 TGTGAATAAGCTTTGACTTATTTTGGAACTTTTAAATGTTTCCAAATACCTATTTT 243  
QY 2052 CAAGGGTATCTGAGATGTTTAACTTGGTATTTAGAAATTTAAACCTTATTAATTT 2111  
DB 242 CAAGGGTATCTGAGATGTTTAACTTGGTATTTAGAAATTTAAACCTTATTAATTT 183  
QY 2112 TTTCTATCAAR 2122  
DB 182 TTTCTATCCAA 172

## RESULT 4

AAAL26538/c  
ID AAAL26538 standard; cDNA; 2290 BP.

XX AAAL26538;

XX 07-DEC-2001 (first entry)

XX

DE Human breast cancer expressed polynucleotide 18995.

KW Human; breast cancer; cell marker; cytosstatic; ss.

XX OS Homo sapiens.

XX WO200151628-A2.

XX 19-JUL-2001.

XX 10-JAN-2001; 2001WO-US00798.

XX 14-JAN-2000; 2000US-0176077.

XX 14-MAR-2000; 2000US-0189167.

XX 24-MAR-2000; 2000US-0192099.

XX 29-MAR-2000; 2000US-0193480.

XX 15-MAY-2000; 2000US-0205230.

XX 09-JUN-2000; 2000US-0211315.

XX 25-JUL-2000; 2000US-0220534.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Lillie J, Xu Y, Wang Y, Steinmann K;

XX WPI; 2001-451856/48.

XX New peptide useful as a marker for the diagnosis of breast cancer -  
XX Claim 1; Page 3544-3545; 3695pp; English.  
XX The invention relates to human breast cancer expressed polynucleotides  
XX (AA07544-AA026789) and methods of assessing whether a patient is  
XX afflicted with breast cancer by examining the correlation between the  
XX expression of certain markers and the cancerous state of breast cells.  
XX The polynucleotides and encoded polypeptides are potential markers for  
XX detecting, diagnosing, monitoring, characterizing treating and  
XX potentially preventing breast cancer. The polynucleotides and encoded  
XX polypeptides are also useful for isolating compounds with cytostatic  
XX activity.

XX Sequence 2290 BP; 508 A; 676 C; 572 G; 528 T; 6 other;

XX Query Match 73.0%; Score 2081.8; DB 22; Length 2290;

XX Best Local Similarity 99.7%; Pred. No. 0;

XX Matches 2105; Conservative 1; Mismatches 3; Indels 2; Gaps 2;

QY 12 CCACTATAACGGCGCGCAGTGTCTGGAATTCGCCCTTGCGCAGATCGCTGGCTGCAGTT 71

DB 2280 CCACTATAACGGCGCGC- GTGTGTGGAATTCGCCCTTGCGCAGATCGCTGGCTGCAGTT 2222

QY 72 GCGCGGCGCATGTGGGGCGGCACGCGCGCGCGCTGCCCGGGAACTGCCGCGCGGC 131

DB 2221 GCGCGGCGCATGTGGCGGCACGCGCGCGCGCTGCCCGGGAACTGCCGCGCGGC 2162

QY 132 CGGAGAGCGCTGTGTGTCTCTGGCGTACTGGCGTTGGCGGGCTGGGCTCGGCTG 191

DB 2161 CGGAGAGCGCTGTGTGTCTCTGGCGTACTGGCGTTGGCGGGCTGGGCTCGGCTG 2102

QY 192 CGGCGGAGCGTGGGGCGGGCGGGCTGCCAGCGCGGACCCCGCGCACCCCGCGC 251

DB 2101 CGGCGGAGCGTGGGGCGGGCGGGCTGCCAGCGCGGACCCCGCGCACCCCGCGC 2042

QY 252 CCGGGCGGGCGAGCGGCTCATGCCCGCGCGCGCTGCCCGGAACTGGGCGCG 311

DB 2041 CCGGGCGGGCGAGCGGCTCATGCCCGCGCGCGCTGCCCGGAACTGGGCGCG 1982

QY 312 CCGGGCGGGCGGTGGGCTGACGTGAGGGCGAGAGTGGCGCTGCGAGGAGAGC 371

DB 1981 CCGGGCGGGCGGTGGGCTGACGTGAGGGCGAGAGTGGCGCTGCGAGGAGAGC 1922

QY 372 GTGGGCTGACCAAGATTACATCTACTAGCGACCGCATCTACTGACCCCGCGCTG 431

DB 1921 GTGGGCTGACCAAGATTACATCTACTAGCGACCGCATCTACTGACCCCGCGCTG 1862

QY 432 CCCAGCGCTGGAAACCGCTGTGCAAGAGAAGAAATATGATTATGATAATTTGCCCAGG 491

DB 1861 CCCAGCGCTGGAAACCGCTGTGCAAGAGAAGAAATATGATTATGATAATTTGCCCAGG 1802

QY 492 ACATCTGTTATCATAGCATTTTATAATGAAGCCTGGTCAACTCTCCTTCGACAGTTTAC 551

DB 1801 ACATCTGTTATCATAGCATTTTATAATGAAGCCTGGTCAACTCTCCTTCGACAGTTTAC 1742

QY 552 AGTGTCTCTGAGACATCCCGGATATCTCTAGAGAAGTGTATCTCTGTTAGATGACTAC 611

DB 1741 AGTGTCTCTGAGACATCCCGGATATCTCTAGAGAAGTGTATCTCTGTTAGATGACTAC 1682

QY 612 AGTGTATAGAGACCATGAAAGGAGCGCTTGGCCAAATGAGCTTTTCGGGACTGCCCAAGGTG 671

DB 1681 AGTGTATAGAGACCATGAAAGGAGCGCTTGGCCAAATGAGCTTTTCGGGACTGCCCAAGGTG 1622

QY 672 CGCTGATCCCGCCCAACAAGAGAGAGGCGCTGGTGGAGCCCGGCTGTGGGGCGCTCT 731

DB 1621 CGCTGATCCCGCCCAACAAGAGAGAGGCGCTGGTGGAGCCCGGCTGTGGGGCGCTCT 1562

QY 732 GCGCGAGGGGCGATGTTCTGACCTTCTGACCTGTGACCTGTGAGTCCACGAAGGCTGG 791

DB 1561 GCGCGAGGGGCGATGTTCTGACCTTCTGACCTGTGAGTCCACGAAGGCTGG 1502

QY 792 CTGAGCGCTGTGTCAGAGGATCCATGAAGAGAGTTCGCGAGTGTGTGCCCGGTGATT 851

DB 1501 CTGAGCGCTGTGTCAGAGGATCCATGAAGAGAGTTCGCGAGTGTGTGCCCGGTGATT 1442

QY 852 GATGTGATGACCTGGAACACCTTCGAATACCTGGGGAACCTCCGGGAGACCCAGATCGGC 911

DB 1441 GATGTGATGACCTGGAACACCTTCGAATACCTGGGGAACCTCCGGGAGACCCAGATCGGC 1382

QY 912 GGTTCGACTGGAGGCTGGTGTTCAGTGCACACAGTTCCTCAGAGGAGAGATACGG 971

DB 1381 GGTTCGACTGGAGGCTGGTGTTCAGTGCACACAGTTCCTCAGAGGAGAGATACGG 1322

QY 972 ATGCAATCCCCCGCTCGATGTGATCAGTCTCCAACTATGCTGTGGGCTGTTTGTGTG 1031

DB 1321 ATGCAATCCCCCGCTCGATGTGATCAGTCTCCAACTATGCTGTGGGCTGTTTGTGTG 1262

QY 1032 AGTAAGAAATATTTGAAATCTGGGCTCTTATGATACAGAAATGGAAGTTTGGGAGGA 1091

DB 1261 AGTAAGAAATATTTGAAATCTGGGCTCTTATGATACAGAAATGGAAGTTTGGGAGGA 1202

QY 1092 GAAACCTCGAATTTCTTTAGATCTGCGAGTGTGGTGGGTTCTGGAACACACCCA 1151

DB 1201 GAAACCTCGAATTTCTTTAGATCTGCGAGTGTGGTGGGTTCTGGAACACACCCA 1142

QY 1152 TGTTCCTCATGTTGGCCATGTTTTCCCAAGAGCTCCCTACTCCCGCAACAAAGGCTCTG 1211

DB 1141 TGTTCCTCATGTTGGCCATGTTTTCCCAAGAGCTCCCTACTCCCGCAACAAAGGCTCTG 1082

QY 1212 GCCAACAGTGTTCGTGAGCTGAAGTATGATGATGATGATGATGATGATGATGATGATGAT 1271

DB 1081 GCCAACAGTGTTCGTGAGCTGAAGTATGATGATGATGATGATGATGATGATGATGATGAT 1023

QY 1272 CGCAACCCCGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGG 1331

DB 1022 CGCAACCCCGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGG 963

QY 1332 GACAAGCTCCAGTGTAAAGACTTCAAGTGGTCTTGGAGACTGTGATCCAGAACTGCAT 1391

DB 962 GACAAGCTCCAGTGTAAAGACTTCAAGTGGTCTTGGAGACTGTGATCCAGAACTGCAT 903

QY 1392 GTGCTGAGAGAGGCTGCTCTTCGGGATGCTCCAGAAACAAAGGACTAACAGACTAC 1451

DB 902 GTGCTGAGAGAGGCTGCTCTTCGGGATGCTCCAGAAACAAAGGACTAACAGACTAC 843

QY 1452 TGCTTTGACTATAACCTCCCGATGAAACCAAGTGTGGGACACCAAGGTCATTCTGTAC 1511

DB 842 TGCTTTGACTATAACCTCCCGATGAAACCAAGTGTGGGACACCAAGGTCATTCTGTAC 783

Qy 1512 CTCTGTGTCATGGATGGCGCAGAAATCAGTTTTCGAGTACAGTCCCGCAGAAAGAAATACGC 1571  
Db |||||||  
782 CTCTGTGTCATGGATGGCGCAGAAATCAGTTTTCGAGTACAGTCCCGCAGAAAGAAATACGC 723  
Qy 1572 TATTAACACCCACAGCCTGAGGGCTGCAATGCTGTGGAAGCAGGAATGATACCTTATC 1631  
Db |||||||  
722 TATTAACACCCACAGCCTGAGGGCTGCAATGCTGTGGAAGCAGGAATGATACCTTATC 663  
Qy 1632 ATGCATCTCTCGGAGAACTGCCCGAGAGATCAGAGTTTCATCTTGCAGGAGATGGA 1691  
Db |||||||  
662 ATGCATCTCTCGGAGAACTGCCCGAGAGATCAGAGTTTCATCTTGCAGGAGATGGA 603  
Qy 1692 TCTTTATTTACGACAGTCCAGAAATGTCTCAGGCTGCGAGGAGGAGTCCAGTGCAC 1751  
Db |||||||  
602 TCTTTATTTACGACAGTCCAGAAATGTCTCAGGCTGCGAGGAGGAGTCCAGTGCAC 543  
Qy 1752 AGTTTGGTTCACCTCTTACGAGACTGCACCACTCGGATCATCAGAAATGTTCTTCAAA 1811  
Db |||||||  
542 AGTTTGGTTCACCTCTTACGAGACTGCACCACTCGGATCATCAGAAATGTTCTTCAAA 483  
Qy 1812 GAGCGCATGTTATGAGCCTGCTGTATCAAGGAGCCCATCGAAGGAGACTGTGAGCCAG 1871  
Db |||||||  
482 GAGCGCATGTTATGAGCCTGCTGTATCAAGGAGCCCATCGAAGGAGACTGTGAGCCAG 423  
Qy 1872 GACTCTGCCCAACAAAGACTTAGCTTAAGCAGTGCACCAACCCCAAACTAGGCTGC 1931  
Db |||||||  
422 GACTCTGCCCAACAAAGACTTAGCTTAAGCAGTGCACCAACCCCAAACTAGGCTGC 363  
Qy 1932 ATTGCTTTGAGAGGCAATCATTTTGGCATTGTTGAAAGTTGTTGGATTAGTAAAAA 1991  
Db |||||||  
362 ATTGCTTTGAGAGGCAATCATTTTGGCATTGTTGAAAGTTGTTGGATTAGTAAAAA 303  
Qy 1992 TGTGAATAAGCTTTGCTACTTATTTTGAACTTTTAAATGTTCCAAATACCTTATTTT 2051  
Db |||||||  
302 TGTGAATAAGCTTTGCTACTTATTTTGAACTTTTAAATGTTCCAAATACCTTATTTT 243  
Qy 2052 CAAGGGTAAATCGTAAGATGTTAAACCTTTGCTATTAGAAAAATTAACCTTATTAATTT 2111  
Db |||||||  
242 CAAGGGTAAATCGTAAGATGTTAAACCTTTGCTATTAGAAAAATTAACCTTATTAATTT 183  
Qy 2112 TTTCTATCAAR 2122  
Db |||||||  
182 TTTCTATCAAA 172

## RESULT 5

AAC98209  
ID AAC98209 standard; cDNA; 1259 BP.

AC AAC98209;

DT 09-MAR-2001 (first entry)

DE Human colon cancer antigen nucleotide sequence SEQ ID NO:219.

KW Human; colon cancer; colon cancer antigen; diagnosis; detection;  
KW identification; cytostatic; cardioactive; neuroprotective; vulnary;  
KW immunomodulatory; muscular; gynaecological; gastrointestinal;  
KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;  
KW neural disorder; immune system disorder; muscular disorder;  
KW reproductive disorder; gastrointestinal disorder; renal disorder;  
KW infectious disease; cardiovascular disorder; ss.

OS Homo sapiens.

FN WO200055351-A1.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US05883.

XX 12-MAR-1999; 99US-0124270.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Ruben SM;

WPI; 2000-587534/55.

P-PSDB; AAB53452.

Colon cancer associated gene sequences, referred to as colon cancer  
antigens, useful for the treatment, prevention, and diagnosis of colon  
disorders such as colon cancer -

Claim 1; Page 642; 2104pp; English.

AAC97991 to AAC98763 encode the human colon cancer associated proteins,  
called human colon cancer antigens, given in AAB53234 to AAB54006. The  
human colon cancer antigens can have cytostatic, cardioactive, muscular;  
neuroprotective, immunomodulatory, gynaecological, gastrointestinal, and  
vulnerable, nephrotropic, antiinfective and antibacterial activities, and  
can be used in gene therapy. The colon cancer antigen polynucleotides,  
proteins and antibodies to the proteins are useful for the prevention,  
treatment and diagnosis of colon disorders, such as colon cancer. The  
polynucleotides may be used in diagnostics and research, such as for  
chromosome identification, and as hybridisation probes. The proteins  
may also be used to prevent diseases such as neural disorders, immune  
system disorders, muscular disorders, reproductive disorders,  
gastrointestinal disorders, wounds, renal disorders, infectious  
diseases, and cardiovascular disorders. AAC98764 to AAC98772 and  
AAB54007 represent sequences used in the exemplification of the present  
invention.

Sequence 1259 BP; 284 A; 310 C; 383 G; 278 T; 4 other;

Query Match 42.4%; Score 1208.4; DB 21; Length 1259;

Best Local Similarity 98.4%; Pred. No. 2.7e-210;  
Matches 1226; Conservative 4; Mismatches 15; Indels 1; Gaps 1;

Qy 292 CGGCGAACCGCGCTGGGCGCGGGGCGAGGCGGTGCGGCTGCGAGCTGCAGGCGGAGGAGC 351

Db 7 CGGTGCGGAAACCGCTGGGCGGGGAGCGAGGCGGTGCGGCTGCGAGCTGCAGGCGGAGGAGC 66

Qy 352 TGGGCTGTCAGGAGGAGAGCGTGGGCTGCAACAGATTAACTACCTCAGGCGACCGCA 411

Db 67 TGGGCTGTCAGGAGGAGAGCGTGGGCTGCAACAGATTAACTACCTCAGGCGACCGCA 126

Qy 412 TCTCAGTGCACCGCGCTGCGGCGCGGCGGTGCGGCTGCGAGCTGCAGGCGGAGGAGTATG 471

Db 127 TCTCAGTGCACCGCGCTGCGGCGCGGCGGTGCGGCTGCGAGCTGCAGGCGGAGGAGTATG 186

Qy 472 ATTATGATAATTTGCCCGAGGACATCTGTTATCATAGCATTTTATAATGAAGCCTGGTCAA 531

Db 187 ATTATGATAATTTGCCCGAGGACATCTGTTATCATAGCATTTTATAATGAAGCCTGGTCAA 246

Qy 532 CTCTCTTCCGACAGTTTACAGTGTCTTGAGACATCCCGGATATCTCTGTAGAGAGAG 591

Db 247 CTCTCTTCCGACAGTTTACAGTGTCTTGAGACATCCCGGATATCTCTGTAGAGAGAG 306

Qy 592 TGATCTCTTGTAGTACTACAGTATAGAGACACCTGAGGAGCGCTTGGCCAAATGAGC 651

Db 307 TGATCTCTTGTAGTACTACAGTATAGAGACACCTGAGGAGCGCTTGGCCAAATGAGC 366

Qy 652 TTTTCGGGACTGCCCAAGTGGCTGATCCGCGCCAAACAAGAGAGAGGCGCTGGTGGAG 711

Db 367 TTTTCGGGACTGCCCAAGTGGCTGATCCGCGCCAAACAAGAGAGAGGCGCTGGTGGAG 426

Qy 712 CCGGCTGCTGGGGCGCTCTGCGGCGAGGGCGGATGTTCTGACCTTCTGACTGTCTACT 771

Db 427 CCGGCTGCTGGGGCGCTCTGCGGCGAGGGCGGATGTTCTGACCTTCTGACTGTCTACT 486

Qy 772 GTGAGTCCACGAGGGTGGTGGAGCGCTGCTGTCAGAGGATCCATGAGAGAGGTCGG 831

Db 487 GTGAGTCCACGAGGGTGGTGGAGCGCTGCTGTCAGAGGATCCATGAGAGAGGTCGG 545

Qy 832 CAGTGGTGTGCGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 891



Qy 1529 CCAGATCATGTTTTCAGTACAGTCCAGAAAGAAATAGCTATACACCCACCCAGCC 1588  
Db 765 CCAGATCATGTTTTCAGTACAGTCCAGAAAGAAATAGCTATACACCCACCCAGCC 706  
Qy 1589 TGAGGGTGCATTTCTGTGGAAGCAGGAATGATACCTTTATCATGATCTCTGCGAAGA 1648  
Db 705 TGAGGGTGCATTTCTGTGGAAGCAGGAATGATACCTTTATCATGATCTCTGCGAAGA 646  
Qy 1649 AACTGCCCCAGAGATCAGAGTTTCATCTTGCAGGAGTGGATCTTTATTTACAGAAC 1708  
Db 645 AACTGCCCCAGAGATCAGAGTTTCATCTTGCAGGAGTGGATCTTTATTTACAGAAC 586  
Qy 1709 GTCCAAGAAATGTCTCAGGCTGCGAGGAGGAGTCTGAGTACACAGTTTCGTTCCACATCTT 1768  
Db 585 GTCCAAGAAATGTCTCAGGCTGCGAGGAGGAGTCTGAGTACACAGTTTCGTTCCACATCTT 526  
Qy 1769 ACGAGACTGCACCACTCGGATCATCAGAAATGTTCTTCAAGAGCGCATGTTATGAAG 1828  
Db 525 ACGAGACTGCACCACTCGGATCATCAGAAATGTTCTTCAAGAGCGCATGTTATGAAG 466  
Qy 1829 CCTCGTGTATCAAGAGCCCATCGAAGGAGACTGTGGAGCCAGGACTCTGCCCAACAAAG 1888  
Db 465 CCTCGTGTATCAAGAGCCCATCGAAGGAGACTGTGGAGCCAGGACTCTGCCCAACAAAG 406  
Qy 1889 ACTTAGCTTAAGCAGTGCAGAACCCACCAAACTAGGCTGCATTCGTTTGAAGAGGCA 1948  
Db 405 ACTTAGCTTAAGCAGTGCAGAACCCACCAAACTAGGCTGCATTCGTTTGAAGAGGCA 346  
Qy 1949 ATCAATTTGCAATTTGTAAGTGTGTTGATTTAGTAAATGTAAGTGTGTA 2008  
Db 345 ATCAATTTGCAATTTGTAAGTGTGTTGATTTAGTAAATGTAAGTGTGTA 286  
Qy 2009 CTTATTTTGAACCTTTTAAATCTCCAAATACCTTATTTTCAAGGGTAAATCGTAAG 2068  
Db 285 CTTATTTTGAACCTTTTAAATCTCCAAATACCTTATTTTCAAGGGTAAATCGTAAG 226  
Qy 2069 ATGTTAACCTTGTGTTAGAAATTTAAACCTTATTAATTTTCTATCAAR 2122  
Db 225 ATGTTAACCTTGTGTTAGAAATTTAAACCTTATTAATTTTCTATCAAR 172

## RESULT 7

ABV29323/c  
ID ABV29323 standard; cDNA; 1352 BP.  
XX AC ABV29323;  
XX DT 16-SEP-2002 (first entry)  
XX DE Human prostate expression marker cDNA 29314.  
XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
XX KW pharmacogenomic marker; gene; ss.  
XX OS Homo sapiens.  
XX PN WO200160860-A2.  
XX PD 23-AUG-2001.  
XX PF 20-FEB-2001; 2001WO-US05171.  
XX PR 17-FEB-2000; 2000US-183319P.  
XX PR 16-MAR-2000; 2000US-189862P.  
XX PR 25-MAY-2000; 2000US-207454P.  
XX PR 09-JUN-2000; 2000US-211314P.  
XX PR 18-JUL-2000; 2000US-219007P.  
XX PR 13-DEC-2000; 2000US-255281P.

(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

Schlegel R, Endege WO, Monahan JE;

DR WPI; 2001-662795/76.  
XX Novel isolated nucleic acid molecule associated with cancerous state of  
PT prostate cells and correlating with presence of prostate cancer, useful  
PT for detecting presence of prostate cancer, stage of prostate cancer -  
XX Claim 1; Page 6275; 11750pp; English.  
XX The invention relates to an isolated nucleic acid molecule (I) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
CC specification or its complement. (I) is useful for:  
CC (a) assessing whether a patient is afflicted with prostate cancer;  
CC (b) monitoring the progression of prostate cancer in a patient;  
CC (c) assessing the efficacy of a test compound to inhibit prostate  
CC cancer in a patient;  
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
CC in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) determining whether prostate cancer has metastasized in a patient;  
CC (g) determining whether prostate cancer has metastasized in a patient;  
CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
CC patient;  
CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.  
XX  
SQ Sequence 1352 BP; 340 A; 335 C; 299 G; 368 T; 10 other;

Query Match 39.1%; Score 1115.2; DB 23; Length 1352;  
Best Local Similarity 99.5%; Pred. No. 2.4e-193;  
Matches 1128; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

Qy 989 TGTATCATGAGTCTCCAAACAAATGCTGTGGGCTGTTTGTGTGTGAGTAAGAAATTTTGA 1048  
Db 1304 TGCCATGTGTCTCCAAACAAATGCTGTGGGCTGTTTGTGTGTGAGTAAGAAATTTTGA 1245  
Qy 1049 ATATCTGGGGTCTTATGATACAGGAATGGAAGTTTGGGAGGAGAGAAACCTCGAATTTTC 1108  
Db 1244 ATATCTGGGGTCTTATGATACAGGAATGGAAGTTTGGGAGGAGAGAAACCTCGAATTTTC 1185  
Qy 1109 CTTTAGATCTGGCAGTGTGTGGGTTCTGGAAACACACACCCATGTTCCCATGTTGGCCA 1168  
Db 1184 CTTTAGATCTGGCAGTGTGTGGGTTCTGGAAACACACACCCATGTTCCCATGTTGGCCA 1125  
Qy 1169 TGTGTTTCCCAAGCAAGCTCCCTACTCCCGCAACAGGCTCTGGCCAAACAGTGTTCGTGC 1228  
Db 1124 TGTGTTTCCCAAGCAAGCTCCCTACTCCCGCAACAGGCTCTGGCCAAACAGTGTTCGTGC 1066  
Qy 1229 AGCTGAAGTATGGATGGAATTTAAAGAGTCTACTACCATCGCAACCCCGTGGCCG 1288  
Db 1065 AGCTGAAGTATGGATGGAATTTAAAGAGTCTACTACCATCGCAACCCCGTGGCCG 1006  
Qy 1289 CTTTGGACCTTTTGGGATGTGACAGAGAGGAGCAGCTCCGGAGCAAGCTCCAGTGTAA 1348  
Db 1005 CTTTGGACCTTTTGGGATGTGACAGAGAGGAGCAGCTCCGGAGCAAGCTCCAGTGTAA 946  
Qy 1349 AGACTTCAAGTGTGTTCTTGAGAGTGTGTATCCAGAACTGCATGTGCTTGAGGACAGGCC 1408  
Db 945 AGACTTCAAGTGTGTTCTTGAGAGTGTGTATCCAGAACTGCATGTGCTTGAGGACAGGCC 886  
Qy 1409 TGGCTTCTTCGGATGCTCCAGAACAAAGGACTATGCTTTGCTATGCTATATACCC 1468  
Db 885 TGGCTTCTTCGGATGCTCCAGAACAAAGGACTATGCTTTGCTATGCTATATACCC 826  
Qy 1469 TCCCGATGAAGAACAGATTTGGGACACAGGCTCATTTCTGCTCTGCTATGGGATGGG 1528  
Db 825 TCCCGATGAAGAACAGATTTGGGACACAGGCTCATTTCTGTACCTCTGTCTATGGGATGGG 766  
Qy 1529 CCAGAAATCAGTGTTCGAGTACAGTCCAGAAAGAAATACGCTATATACACCCACCCAGCC 1588  
Db 765 CCAGAAATCAGTGTTCGAGTACAGTCCAGAAAGAAATACGCTATATACACCCACCCAGCC 706  
Qy 1589 TGAGGGTGCATTTGCTGTGGAAGCAGGAATGGATACCCCTATCATGATCTCTCGGAAGA 1648  
Db 705 TGAGGGTGCATTTGCTGTGGAAGCAGGAATGGATACCCCTATCATGATCTCTCGGAAGA 646

QY 1649 AACTGCCCCAGAGAAATCAGAAATTCATCTCTGAGGAGATGGATCTTTATTTACGAAACA 1708  
DB 645 AACTGCCCCAGAGAAATCAGAAATTCATCTCTGAGGAGATGGATCTTTATTTACGAAACA 586  
QY 1709 GTCCAGAAATGTCTCAGGCTCGAGGAGGAGTGCAGTGCACAGTTTCCTCCACTCTT 1768  
DB 585 GTCCAGAAATGTCTCAGGCTCGAGGAGGAGTGCAGTGCACAGTTTCCTCCACTCTT 526  
QY 1769 ACAGAGACTGCACCAACTCGGATCATCAGAAATGTTCTTCAAGAGCGCATGTTATGAAG 1828  
DB 525 ACAGAGACTGCACCAACTCGGATCATCAGAAATGTTCTTCAAGAGCGCATGTTATGAAG 466  
QY 1829 CTTGCTGTATCAAGAGGCCCATCGAAGGAGACTGTGAGCGAGACTCTGCCCAACAAAG 1888  
DB 465 CTTGCTGTATCAAGAGGCCCATCGAAGGAGACTGTGAGCGAGACTCTGCCCAACAAAG 406  
QY 1889 ACTTAGCTAAGCAGTACACCAAGACCCACCAAACTAGGCTGCATTCCTTGAAGGCA 1948  
DB 405 ACTTAGCTAAGCAGTACACCAAGACCCACCAAACTAGGCTGCATTCCTTGAAGGCA 346  
QY 1949 ATCATTTTGCCTTTGTGAAAGTTGTGTTGATTTAGTAAATTTAGTAAATTTAGTAA 2008  
DB 345 ATCATTTTGCCTTTGTGAAAGTTGTGTTGATTTAGTAAATTTAGTAAATTTAGTAA 286  
QY 2009 CTTATTTTGAGAACTTTTAAATGTTTCCAAATACCTTATTTTCAAGGGTAAATCGTAAG 2068  
DB 285 CTTATTTTGAGAACTTTTAAATGTTTCCAAATACCTTATTTTCAAGGGTAAATCGTAAG 226  
QY 2069 ATGTTAACCTCTGGTATTTAGAAATTTAAACCTTATATATTTTCTATCAAR 2122  
DB 225 ATGTTAACCTCTGGTATTTAGAAATTTAAACCTTATATATTTTCTATCAAR 172

## RESULT 8

AAH33532  
ID AAH33532 standard; cDNA; 1517 BP.

XX AC AAH33532;

DT 03-SEP-2001 (first entry)

XX DE Human colon cancer antigen encoding cDNA SEQ ID NO:588.

XX KW Human; colon cancer; colon cancer antigen; diagnosis; detection;

XX KW colorectal carcinoma; ss.

XX OS Homo sapiens.

XX PN WO20012920-A2.

XX PN 05-APR-2001.

XX PD 28-SEP-2000; 2000WO-US26524.

XX PF 29-SEP-1999; 99US-0157137.

XX PR 03-NOV-1999; 99US-0163280.

XX XX (HUMA-) HUMAN GENOME SCI INC.

XX PI Ruben SM, Barash SC, Birse CE, Rosen CA;

XX DR WPI; 2001-235357/24.

XX DR P-PSDB; AAG74101.

XX PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
PT useful for preventing, diagnosing and/or treating colorectal cancers -

XX PS Claim 1; Page 2640; 9803pp; English.

XX CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon  
CC cancer-associated nucleic acid molecules (N) and proteins (P), where  
CC the proteins are collectively known as colon cancer antigens. The colon

CC cancer antigens have cytostatic activity and can be used in gene  
CC therapy and vaccine production. N and P may be used in the prevention,  
CC diagnosis and treatment of diseases associated with inappropriate P  
CC expression. For example, N and P may be used to treat disorders  
CC associated with decreased expression by rectifying mutations or deletions  
CC in a patient's genome that affect the activity of P by expressing  
CC inactive proteins or to supplement the patients own production of P.  
CC Additionally, N may be used to produce the colon cancer-associated Ps,  
CC by inserting the nucleic acids into a host cell and culturing the cell  
CC to express the proteins. N and P can be used in the prevention, diagnosis  
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204  
CC and AAB77789 represent sequences used in the exemplification of the  
CC present invention.

CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were  
CC missing at time of publication, meaning no sequences are present for  
CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX SQ Sequence 1517 BP; 366 A; 360 C; 363 G; 425 T; 3 other;

Query Match 35.5%; Score 1012.2; DB 22; Length 1517;  
Best Local Similarity 99.3%; Pred. No. 1.3e-174;  
Matches 1008; Conservative 7; Mismatches 0; Indels 0; Gaps 0;

QY 1115 GATCTGGCAGTGTGGGTCTCTGAAACACACCCATGTTCCATGTTGGCCATGTTT 1174  
DB 479 GATCTGGCAGTGTGGGTCTCTGAAACACACCCATGTTCCATGTTGGCCATGTTT 538  
QY 1175 CCCCAAGCAAGCTCCCTACTCCCGCAACAAAGGCTCTGGCCAAACAGTCTTCGTCAGCTGA 1234  
DB 539 CCCCAAGCAAGCTCCCTACTCCCGCAACAAAGGCTCTGGCCAAACAGTCTTCGTCAGCTGA 598  
QY 1235 AGTATGGATGGATGAATTTAAAGAGCTCTACTACCATCGCAACCCCGTCCCGCTTGA 1294  
DB 599 AGTATGGATGGATGAATTTAAAGAGCTCTACTACCATCGCAACCCCGTCCCGCTTGA 658  
QY 1295 ACCTTTTGGGATGTGACAGAGAGAGAGAGCTCCGGGCAACAGCTCCAGTGTAAAGACTT 1354  
DB 659 ACCTTTTGGGATGTGACAGAGAGAGAGAGCTCCGGGCAACAGCTCCAGTGTAAAGACTT 718  
QY 1355 CAAGTGTCTTGGAGACTGTGTATCCAGAACTGCTGCTGAGGACAGGCTGGCTT 1414  
DB 719 CAAGTGTCTTGGAGACTGTGTATCCAGAACTGCTGCTGAGGACAGGCTGGCTT 778  
QY 1415 CTTTCGGGATGCTCCAGAAACAAAGACTAACAGACTACTGCTTTGACTATACCTCCCGA 1474  
DB 779 CTTTCGGGATGCTCCAGAAACAAAGACTAACAGACTACTGCTTTGACTATACCTCCCGA 838  
QY 1475 TGAACACAGATTTGTGGGACACAGGTCAATCTGTACCTCTGTCTGAGGATGGGCCAGAA 1534  
DB 839 TGAACACAGATTTGTGGGACACAGGTCAATCTGTACCTCTGTCTGAGGATGGGCCAGAA 898  
QY 1535 TCAGTTTTTTCAGTACACGTCCAGAAAGAAATACGCTATAACACCCAGCTGAGGG 1594  
DB 899 TCAGTTTTTTCAGTACACGTCCAGAAAGAAATACGCTATAACACCCAGCTGAGGG 958  
QY 1595 CTGCATTGCTGTGGAAGCAGGAATGATACCTTATCATGCTCTCTCGAAGAACTGC 1654  
DB 959 CTGCATTGCTGTGGAAGCAGGAATGATACCTTATCATGCTCTCTCGAAGAACTGC 1018  
QY 1655 CCCAGAGAAATCAGAACTTCAATTTGAGGAGATGGAATCTTTATTTACGAAACAGTCAA 1714  
DB 1019 CCCAGAGAAATCAGAACTTCAATTTGAGGAGATGGAATCTTTATTTACGAAACAGTCAA 1078  
QY 1715 GAAATGTCTCCAGGCTCCGAGGAGAGTGCAGTGCAGTTCCTTCCCTCTTACGAGA 1774  
DB 1079 GAAATGTCTCCAGGCTCCGAGGAGAGTGCAGTGCAGTTCCTTCCCTCTTACGAGA 1138  
QY 1775 CTGCACCAACTCGGATCATCAGAAATGGTTCCTTCAAGAGCGCATGTTATGAAGCCCTCGT 1834  
DB 1139 CTGCACCAACTCGGATCATCAGAAATGGTTCCTTCAAGAGCGCATGTTATGAAGCCCTCGT 1198  
QY 1835 GTATCAAGAGGCCCATCGAAGGAGACTGTGGAGCCAGGACTCTGCCCAACAAAGACTTAG 1894



1199 GTATCAAGGAGCCCATCGAAGGAGACTGTGGAGCCAGGACTCTGCCCAACAAAGACTTAG 1258  
1895 CTAGCAGTGCACGAGACCCACCAAAACCTAGGCTGCATTCCTTTGAGAGGCAATCATT 1954  
1259 CTAAGCAGTGCACGAGACCCACCAAAACCTAGGCTGCATTCCTTTGAGAGGCAATCATT 1318  
1955 TTGCGCATTTGTGAAGTTGTGTTGGATTAGTAAATAATGCAATAAGCTTTGTACTTATT 2014  
1319 TTGCCATTGTGAAGTTGTGTTGGATTAGTAAATAATGCAATAAGCTTTGTACTTATT 1378  
2015 TTGAGAACTTTTAAATGTTCCAAATACCCCTATTTTCAAAAGGGTAATCGTAAGATGTTA 2074  
1379 TTGAGAACTTTTAAATGTTCCAAATACCCCTATTTTCAAAAGGGTAATCGTAAGATGTTA 1438  
2075 ACCCTGGTATTAGAAATTAACCTTATATATATTTTCTATCAARAWA 2129  
1439 ACCCTGGTATTAGAAATTAACCTTATATATATTTTCTATCAARAWA 1493

## RESULT 9

ABL90290  
ID - ABL90290 standard; cDNA; 1517 BP.  
XX  
AC ABL90290;  
XX  
DT 24-MAY-2002 (first entry)  
XX  
DE Human polynucleotide SEQ ID NO 852.  
XX  
KW Cytostatic; immunosuppressive; nontropic; neuroprotective; antiviral;  
KW antileptic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;  
KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; human; secreted protein; gene; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200190304-A2.  
XX  
PD 29-NOV-2001.  
XX  
PF 18-MAY-2001; 2001WO-US16450.  
XX  
PR 19-MAY-2000; 2000US-205515P.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.

Birse CE, Rosen CA;  
WPI; 2002-122018/16.  
P-PSDB; ABB99881.  
XX  
XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and  
XX prevention of neural, immune system, muscular, reproductive,  
XX gastrointestinal, pulmonary, cardiovascular, renal and proliferative  
XX disorders -  
XX  
XX Claim 4; SEQ ID NO 852; 2081pp + Sequence Listing; English.

XX  
XX The invention relates to novel genes (ABL89449-ABL90853) and proteins  
XX (ABB89040-ABB90444) useful for preventing, treating or ameliorating  
XX medical conditions e.g. by protein or gene therapy. The genes are  
XX isolated from a range of human tissues disclosed in the specification.  
XX The nucleic acids, proteins, antibodies and (ant)agonists are useful  
XX in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
XX and ovarian cancer and other cancers of the adrenal gland, bone, bone  
XX marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
XX (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
XX haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
XX disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
XX colitis; (c) cardiovascular disorders such as myocardial ischaemia;  
XX (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
XX epilepsy; and (f) infectious diseases such as viral, bacterial, fungal

CC and parasitic infections.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX  
SQ Sequence 1517 BP; 366 A; 360 C; 363 G; 425 T; 3 other;

Query Match 35.5%; Score 1012.2; DB 24; Length 1517;

Best Local Similarity 99.3%; Pred. No. 1.3e-174;

Matches 1008; Conservative 7; Mismatches 0; Indels 0; Gaps 0;

QY 1115 GATCTGGCAGTGTGGGGTCTTGAAACACACCCCATGTTCCCATGTTGGCCATGTTTT 1174

Db 479 GATCTGGCAGTGTGGGGTCTTGAAACACACCCCATGTTCCCATGTTGGCCATGTTTT 538

QY 1175 CCCCAAGCAAGCTCCCTACTCCCGCAACAGGCTCTGGCCACAGTGTTCGTGCAGCTGA 1234

Db 539 CCCCAAGCAAGCTCCCTACTCCCGCAACAGGCTCTGGCCACAGTGTTCGTGCAGCTGA 598

QY 1235 AGTATGGATGGATGAATTTAAAGAGCTTACTACCATCGCAACCCCGTCCCGCTTGGGA 1294

Db 599 AGTATGGATGGATGAATTTAAAGAGCTTACTACCATCGCAACCCCGTCCCGCTTGGGA 658

QY 1295 ACCTTTTGGGGATGTGACAGAGAGGAGCAGCTCCGGGCAAGCTCCAGTGTAAAGACTT 1354

Db 659 ACCTTTTGGGGATGTGACAGAGAGGAGCAGCTCCGGGCAAGCTCCAGTGTAAAGACTT 718

QY 1355 CAAGTGTCTTTGGAGACTGTGTATCCAGACTGTCATGTCCTGAGGACAGGCTGGCTT 1414

Db 719 CAAGTGTCTTTGGAGACTGTGTATCCAGAACTGTCATGTCCTGAGGACAGGCTGGCTT 778

QY 1415 CTTTCGGATGCTCCAGAAACAAAGACTTAACAGACTACTGCTTTGACTATAACCTCCCGA 1474

Db 779 CTTTCGGATGCTCCAGAAACAAAGACTTAACAGACTACTGCTTTGACTATAACCTCCCGA 838

QY 1475 TGAACACCAAGTGTGGGACACAGGTCTTGTACTCTGTCTGTCAGGATGGGCCAGAA 1534

Db 839 TGAACACCAAGTGTGGGACACAGGTCTTGTACTCTGTCTGTCAGGATGGGCCAGAA 898

QY 1535 TCAGTCTTTTCAGTACACGTCCTCCAGAAAGAAATACGCTATAACACCCAGGCTGAGGG 1594

Db 899 TCAGTCTTTTCAGTACACGTCCTCCAGAAAGAAATACGCTATAACACCCAGGCTGAGGG 958

QY 1595 CTGCATTGTGTGGAAGCAGGAATGATACCTTATCATGCTCTCTCGAAGAACTGC 1654

Db 959 CTGCATTGTGTGGAAGCAGGAATGATACCTTATCATGCTCTCTCGAAGAACTGC 1018

QY 1655 CCCAGAGAATCAGAAGTTCTATTGAGGAGGATGGATCTTTATTTCCAGAACTGCTCAA 1714

Db 1019 CCCAGAGAATCAGAAGTTCTATTGAGGAGGATGGATCTTTATTTCCAGAACTGCTCAA 1078

QY 1715 GAAATGTCTCAGGCTCGAGGAGGAGTGCAGTGCAGTTCGTTCCACTCTTACGAGA 1774

Db 1079 GAAATGTCTCAGGCTCGAGGAGGAGTGCAGTGCAGTTCGTTCCACTCTTACGAGA 1138

QY 1775 CTCACCAACTCGGATCATCAGAAATGTTCTTCAAGAGCGCATGTTATGAAGCCCTCGT 1834

Db 1139 CTCACCAACTCGGATCATCAGAAATGTTCTTCAAGAGCGCATGTTATGAAGCCCTCGT 1198

QY 1835 GTATCAAGGAGCCCATCGAAGGAGACTGTGGAGCCAGGACTCTGCCCAACAAAGACTTAG 1894

Db 1199 GTATCAAGGAGCCCATCGAAGGAGACTGTGGAGCCAGGACTCTGCCCAACAAAGACTTAG 1258

QY 1895 CTAAGCAGTGCACGAGACCCACCAAAACCTAGGCTGCATTCCTTTGAGAGGCAATCATT 1954

Db 1259 CTAAGCAGTGCACGAGACCCACCAAAACCTAGGCTGCATTCCTTTGAGAGGCAATCATT 1318

QY 1955 TTGCGCATTTGTGAAGTTGTGTTGGATTAGTAAATAATGCAATAAGCTTTGTACTTATT 2014

Db 1319 TTGCGCATTTGTGAAGTTGTGTTGGATTAGTAAATAATGCAATAAGCTTTGTACTTATT 1378

QY 2015 TTGAGAACTTTTAAATGTTCCAAATACCCCTATTTTCAAAAGGGTAATCGTAAGATGTTA 2074

Db 1439 TTGAGAACTTTTAAATGTTCCAAATACCCCTATTTTCAAAAGGGTAATCGTAAGATGTTA 1493



Db 1379 TTGAGAACTTTTAAATGTTCCAAATACCCCTATTTTCAAGGGTAATCGTAAGATGTTA 1438  
Qy 2075 ACCCTTGGTATTTAGAAAAATTAACCTTAATAATTTTCTATCAARAWRWA 2129  
Db 1439 ACCCTTGGTATTTAGAAAAATTAACCTTAATAATTTTCTATCAARAWRWA 1493

## RESULT 10

AA78414

ID AA78414 standard; cDNA; 1517 BP.

XX AC

XX AA78414;

XX

XX 20-NOV-2000 (first entry)

XX

XX Human secreted protein gene 34 SEQ ID NO:44.

XX

XX Human; secreted protein; cytotatic; antianaemic; antidiabetic;

XX antiinflammatory; ophthalmological; antirheumatic; antithratic;

XX antiporatic; antilangogenic; cardiac; anti-HIV; nootropic;

XX neuroprotective; antimicrobial; antiparkinsonian; cancer;

XX immune system disorder; angiogenesis; hyperproliferative disorder;

XX cardiovascular disorder; apoptosis; neurological disease;

XX infectious disease; wound healing; ss.

XX

XX Homo sapiens.

XX

XX WO200035937-A1.

XX

XX 22-JUN-2000.

XX

XX 16-DEC-1999; 99WO-US299950.

XX

XX 17-DEC-1998; 98US-0112809.

XX

XX 18-DEC-1998; 98US-0113006.

XX

XX (HUMA-) HUMAN GENOME SCI INC.

XX

XX Ruben SM, Ebner R, Rosen CH, Endress GA, Soppet DR, Ni J;

XX Duan DR, Moore PA, Shi Y, Latleur DW, Olsen HS, Florence K;

XX

XX WPI; 2000-431566/37.

XX

XX P-PSDB; AAB24470.

XX

XX Forty seven human nucleic acids encoding secreted proteins, useful in

XX the treatment, prevention and diagnosis of cancers, disorders of the

XX immune system, angiogenesis disorders, neurological diseases and

XX hyperproliferative disorders -

XX

XX Claim 1; Page 465-466; 562pp; English.

XX

XX The polynucleotide sequence given in AAA78381 to AAA78432 encode the

XX human secreted proteins given in AAB24437 to AAB24604. Human secreted

XX proteins have activities based on the tissues and cells the genes are

XX expressed in. Examples of activities include: cytotatic; antianaemic;

XX antidiabetic; antinflammatory; ophthalmological; antirheumatic;

XX antithratic; antiporatic; antilangogenic; cardiac; anti-HIV;

XX nootropic; neuroprotective; antimicrobial and antiparkinsonian.

XX Human secreted protein polynucleotides, polypeptides, antagonists and/or

XX agonists may be useful in treating, preventing, and/or diagnosing other

XX diseases, disorders, and/or conditions such as: (a) cancers; (b)

XX disorders of the immune system; (c) angiogenesis disorders; (d)

XX hyperproliferative disorders; (e) cardiovascular disorders; (f) diseases

XX associated with increase apoptosis; (g) neurological diseases; and

XX (h) infectious diseases. They are also used to promote wound healing.

XX AAA78372 to AAA78380 and AAB24436 represent sequences used in the

XX exemplification of the present invention.

XX

XX Sequence 1517 BP; 366 A; 360 C; 363 G; 424 T; 4 other;

XX

XX Query Match 35.58; Score 1011.8; DB 21; Length 1517;

XX Best Local Similarity 99.24; Pred. No. 1.5e-174;

XX Matches 1007; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

XX

XX

XX

XX

XX

XX

XX



Human; breast cancer; cell marker; cytostatic; ss.  
Homo sapiens.  
WO200151628-A2.  
19-JUL-2001.  
10-JAN-2001; 2001WO-US00798.  
14-JAN-2000; 2000US-0176077.  
14-MAR-2000; 2000US-0189167.  
24-MAR-2000; 2000US-0192099.  
29-MAR-2000; 2000US-0193480.  
15-MAY-2000; 2000US-0205230.  
09-JUN-2000; 2000US-0211315.  
25-JUL-2000; 2000US-0220534.  
(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
Lillie J, Xu Y, Wang Y, Steinmann K;  
WPI; 2001-451856/48.  
New peptide useful as a marker for the diagnosis of breast cancer -  
Claim 1; Page 3592-3593; 3695pp; English.  
The invention relates to human breast cancer expressed polynucleotides (AAL07544-AAL26789) and methods of assessing whether a patient is afflicted with breast cancer by examining the correlation between the expression of certain markers and the cancerous state of breast cells. The polynucleotides and encoded polypeptides are potential markers for detecting, diagnosing, monitoring, characterizing treating and potentially preventing breast cancer. The polynucleotides and encoded polypeptides are also useful for isolating compounds with cytostatic activity.  
Sequence 883 BP; 273 A; 172 C; 134 G; 289 T; 15 other;  
Query Match 26.5%; Score 756.2; DB 22; Length 883;  
Best Local Similarity 98.6%; Pred. No. 3.8e-128;  
Matches 788; Conservative 5; Mismatches 3; Indels 3; Gaps 3;  
QY 2031 TGTTCACAAATACCTATTTTCAAGGGTAATCGTAAGATGTTAAACCCCTGGTATTAGA 2090  
DB 873 TGTTCACAAATACCTATTTTCAAGGGTAATCGTAAGATGTTAAACCC-TGGTATTAGA 815  
QY 2091 AAATTAACCTTATATATATTTTCTATCAARAWRWAWATTTTACAGTCGTGCTTTTAC 2150  
DB 814 AAATTAACCTTATATATATTTTCTATCAAGATGTTATTTACAGTCGTGCTTTTAC 755  
QY 2151 TCTCATTAGCAAAAAGATAAAGATTTTATTTGGTATTATACAGATTTCCAGGTACGA 2210  
DB 754 TCTCATTAGCAAAAAGATAAAGA-TTTATTTGGTATTATACAGAAATTTCCAGGTACGA 696  
QY 2211 AGATATCTGCATGGTGGGAATCAGGTTCAAGCAACCTACTTTGTCATTAACTGATATAC 2270  
DB 695 AGATATCTGCATGGTGGGAATCAGGTTCAAGCAACCTACT-TTGCATTAACTGATATAC 637  
QY 2271 CTGAGTCGGGGTTAAAGTTTCCAGATAGAGACATGTCACCTAGGAACATTTGATT 2330  
DB 636 CTCAGTCGGGGTTAAAGTTTCCAGATAGAGACATGTCACCTAGGAACATTTGATT 577  
QY 2331 GATTATTACAGTCATGAGATCTTCTAGATGTTATTTTAAAGATGCTTTTGGTTAT 2390  
DB 576 GATTATTACAGTCATGAGATCTTCTAGATGTTATTTTAAAGATGCTTTTGGTTAT 517  
QY 2391 GTGTGTGTACACAGTTTAAACCTCCATAATGTTTCATGTCAGCCAAAGAGGACTAACCAAA 2450  
DB 516 GTGTGTGTACACAGTTTAAACCTCCATAATGTTTCATGTCAGCCAAAGAGGACTAACCAAA 457  
QY 2451 GCTGAAATCTCAGAGAACAAATTTGCTTTTACTAAGCTGAGTCAACTTGAGAGCGCACTTCT 2510

DB 456 GCTGAAATCTCAGAGAACAAATTTGCTTTTACTAAGCTGAGTCAACTTGAGAGCGCACTTCT 397  
QY 2511 AACAAATGCCGCACTGTAGTGTGCTGTCTTACCACTATGACTTTTAAACATGTTTATAT 2570  
DB 396 AACAAATGCCGCACTGTAGTGTGCTGTCTTACCACTATGACTTTTAAACATGTTTATAT 337  
QY 2571 CATTTTAAATTTTATGATACGCTAGTGTGTCAGGAGAAATGTAATGTTCTATATGAAAT 2630  
DB 336 CATTTTAAATTTTATGATACGCTAGTGTGTCAGGAGAAATGTAATGTTCTATATGAAAT 277  
QY 2631 CCTTTTCAAGTTTGTTCATTATAACAGTTATTAATTTAAACAGCGTTAGATTGTG 2690  
DB 276 CCTTTTCAAGTTTGTTCATTATAACAGTTATTAATTTAAACAGCGTTAGATTGTG 217  
QY 2691 CTGCTCAACTGCTGTGAAATTTCTCTGAGTAAATTTCTGATTTGTGAATGATCCAGACC 2750  
DB 216 CTGCTCAACTGCTGTGAAATTTCTCTGAGTAAATTTCTGATTTGTGAATGATCCAGACC 157  
QY 2751 AACCTGAGATTTTGTCAACCTGATTAAAGTCAATATGAATGATTAAAAAGATGTGAGAAC 2810  
DB 156 AACCTGAGATTTTGTCAACCTGATTAAAGTCAATATGAATGATTAAAAAGATGTGAGAAC 97  
QY 2811 AAAAAAAAAAAAAAAAAAAAAA 2829  
DB 96 ACTGAAAAAAAAAAAAAAAAAAAA 78  
RESULT 13  
AAL02624  
ID AAL02624 standard; cDNA; 592 BP.  
XX AAL02624;  
XX AC AAL02624;  
DT 19-MAY-2000 (first entry)  
XX Human colon cancer cell line polynucleotide sequence SEQ ID NO:2615.  
DE Human; colon cancer; tumour; diagnosis; gene expression product;  
XX probe; detection; cancerous state; metastasis; identification;  
KW breast cancer; oestrogen receptor-positive breast cancer; therapy;  
KW oestrogen receptor-negative breast cancer; lung cancer; ss.  
XX Homo sapiens.  
OS XX  
PN WO958675-A2.  
XX 18-NOV-1999.  
PD 13-MAY-1999; 99WO-US10602.  
XX PF 14-MAY-1998; 98US-0085426.  
XX PR 15-MAY-1998; 98US-0085537.  
PR 15-MAY-1998; 98US-0085696.  
PR 21-OCT-1998; 98US-0105234.  
PR 27-OCT-1998; 98US-0105877.  
XX (CHIR) CHIRON CORP.  
PA (HYSE-) HYSEQ INC.  
XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;  
PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;  
PI Lamson G, Drmanac R, Cckvenjakov R, Dickson M, Drmanac S, Labat I;  
PI Leshkowitz D, Kita D, Garcia V, Jones LM, Stache-Crain B;  
XX WPI; 2000-126369/11.  
DR Polynucleotide library used to determine cancerous states of mammalian  
XX cells -  
PT Claim 1; Page 1061-1062; 1097pp; English.  
PS AAA00010 to AAA02716 represent polynucleotides isolated from cDNA  
XX

CC libraries constructed from human colon cancer cell lines. The present  
CC invention also describes a method of detecting differentially expressed  
CC genes correlated with a cancerous state of a mammalian cell, comprising  
CC detecting at least one differentially expressed gene product in a test  
CC sample derived from a cell suspected of being cancerous, where detection  
CC of the differentially expressed gene product is correlated with a  
CC cancerous state of the cell from which the test sample was derived.  
CC The polynucleotide sequences can be used in a method for detecting  
CC differentially expressed genes correlated with a cancerous state of a  
CC mammalian cell. The polynucleotides can also be used as probes for  
CC detecting and mapping related genes. They can be used in diagnosis and  
CC prognosis of diseases and disorders (e.g. identification of  
CC pre-metastatic or metastatic cancerous states, stages of cancer, or  
CC responsiveness of cancer to therapy). This is particularly for breast  
CC cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-  
CC negative breast cancer, lung cancer, and colon cancer.  
XX  
SQ Sequence 592 BP; 173 A; 115 C; 122 G; 163 T; 19 other;

Query Match 19.4%; Score 552.8; DB 21; Length 592;  
Best Local Similarity 96.1%; Pred. No. 3.2e-91;  
Matches 565; Conservative 19; Mismatches 1; Indels 3; Gaps 2;

QY 1535 TCAGTGTTCAGTACAC-CGTCCAGAGAAATACGCTATACACCCACGAGCTGAGG 1593  
DB 5 TTAGTGTTCAGTACACCGTCCAGAGAAATACGCTATACACCCACGAGCTGAGG 64  
QY 1594 GCTGCATGTGTGGAAGCAGGAATGATACCCCTTATCATGCTCTCTGGAAGAACTG 1653  
DB 65 GCTGCATGTGTGGAAGCAGGAATGATACCCCTTATCATGCTCTCTGGAAGAACTG 124  
QY 1654 -CCCCAGAGNATCAGAGTTCATCTTCGAGGAGTGGATCTTATTCACGACAGTC 1711  
DB 125 ATSMCYGMSWRTCAKWKMTYATCTTGYWKGAGTGGATCTTATTCACGACAGTC 184  
QY 1712 CAAGAAATGTGTCCAGGCTGCGAGAGAGGAGTGCAGTGCAGATTTTCGTCACCTCTTACG 1771  
DB 185 CAGAAATGTGTCCAGGCTGCGAGAGAGGAGTGCAGTGCAGATTTTCGTCACCTCTTACG 244  
QY 1772 AGATGCACCAACTCGGATCATCAGAAATGTTCTTCAAGAGCGCATGTTATGAAGCCT 1831  
DB 245 AGATGCACCAACTCGGATCATCAGAAATGTTCTTCAAGAGCGCATGTTATGAAGCCT 304  
QY 1832 CGTGTATCAGAGGCCCATCGAAGGAGCTGTGGAGCCGAGCTCTGCCCAACAAAGACT 1891  
DB 305 CGTGTATCAGAGGCCCATCGAAGGAGCTGTGGAGCCGAGCTCTGCCCAACAAAGACT 364  
QY 1892 TAGCTAAGCAGTACACCAAGCCCAACCAAACTAGGCTGCATTTGCTTTGAAGAGGCAATC 1951  
DB 365 TAGCTAAGCAGTACACCAAGCCCAACCAAACTAGGCTGCATTTGCTTTGAAGAGGCAATC 424  
QY 1952 ATTTTGCATTTGTGAAGTTGTGTGGAATTAAGTAAATGTAAGCTTTGTACTT 2011  
DB 425 ATTTTGCATTTGTGAAGTTGTGTGGAATTAAGTAAATGTAAGCTTTGTACTT 484  
QY 2012 ATTTTGAAGCTTTTAAATGTTCCAAATACCTTATTTTCAAGGGTAAATCGTAAGATG 2071  
DB 485 ATTTTGAAGCTTTTAAATGTTCCAAATACCTTATTTTCAAGGGTAAATCGTAAGATG 544  
QY 2072 TTAACCCCTTGTGATTTAGAAAATTAACACCTTATTAATATTTTCTATC 2119  
DB 545 TTAACCCCTTGTGATTTAGAAAATTAACACCTTATTAATATTTTCTATC 592

RESULT 14  
AAH82814  
ID AAH82814 standard; cDNA; 554 BP.  
XX  
AC AAH82814;  
XX

25-SEP-2001 (first entry)

DE Human ovarian tumour associated polynucleotide sequence SEQ ID NO:438.

XX Human; ovarian tumour; ovarian cancer; diagnosis; gene therapy;  
KW immunogenic; vaccine; ss.  
XX

OS Homo sapiens.

PN WO200151513-A2.

XX 19-JUL-2001.

PD 16-JAN-2001; 2001WO-US01575.

PF 14-JAN-2000; 2000US-0176722.

XX (CORI-) CORIXA CORP.

XX Algate PA;

PI WPI; 2001-425866/45.

XX Novel ovarian tumor proteins, and nucleic acids encoding them, used to  
PT treat and diagnose cancers, particularly ovarian cancer -

XX Claim 5; Page 145; 338pp; English.

XX AAH82377 to AAH83878 represent human ovarian tumour-associated  
CC polynucleotide sequences which encode ovarian tumour proteins. The  
CC ovarian tumour protein and polynucleotide sequences have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. The  
CC ovarian tumour proteins and polynucleotides can be used to inhibit  
CC the development of cancer, particularly ovarian cancer. They can also  
CC be used to diagnose the onset and progression of cancer.

XX Sequence 554 BP; 160 A; 138 C; 135 G; 120 T; 1 other;

Query Match 18.5%; Score 527; DB 22; Length 554;  
Best Local Similarity 98.9%; Pred. No. 1.5e-86;  
Matches 541; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 1400 GGACAGCCCTGGCTTCTTCGGGATGCTCCAGAACAAAGGACTAACAGACTACTGCTTTGA 1459  
DB 4 GGCAGGTCTCTGGCTTCTTCGGGATGCTCCAGAACAAAGGACTAACAGACTACTGCTTTGA 63

QY 1460 CTATAACCCCTCCCGATGAAACCCAGATTTGGGACACACAGGTCACTTCTGTACCTCTGTCA 1519

DB 64 CTATAACCCCTCCCGATGAAACCCAGATTTGGGACACACAGGTCACTTCTGTACCTCTGTCA 123

QY 1520 TGGGATGGGCCAGAAATCAGTTTTTCGAGTACACGTCCAGAAAGAAATAGCTATATACAC 1579

DB 124 TGGGATGGGCCAGAAATCAGTTTTTCGAGTACACGTCCAGAAAGAAATAGCTATATACAC 183

QY 1580 CCACAGCCCTGAGGGCTGCAATTCCTGTGGAAGCAGGATCGATACCCCTTATCATGCACT 1639

DB 184 CCACAGCCCTGAGGGCTGCAATTCCTGTGGAAGCAGGATCGATACCCCTTATCATGCACT 243

QY 1640 CTGCGAAGAACTGCCCCAGAGAAATCAGAAATTCATCTTCAGGAGGATGGATCTTTATT 1699

DB 244 CTGCGAAGAACTGCCCCAGAGAAATCAGAAATTCATCTTCAGGAGGATGGATCTTTATT 303

QY 1700 TCACGAAACAGTCCAAAGAAATGTGTCCAGGTGCGAGGAAGAGTCGAGTGACAGTTTCGT 1759

DB 304 TCACGAAACAGTCCAAAGAAATGTGTCCAGGTGCGAGGAAGAGTCGAGTGACAGTTTCGT 363

QY 1760 TCACCTCTTACGAGATGTCACCAACTCGGATCATCAGAAATGTTCTTCAAGAGGCCAT 1819

DB 364 TCACCTCTTACGAGATGTCACCAACTCGGATCATCAGAAATGTTCTTCAAGAGGCCAT 423

QY 1820 GTTATGAAGCTCTGTGTATCAAGGAGCCCATCGAAGGAGACTGTGGAGCCAGGACTCTGC 1879

DB 424 GTTATGAAGCTCTGTGTATCAAGGAGCCCATCGAAGGAGACTGTGGAGCCAGGACTCTGC 483

QY 1880 CCACAAAGACTTAGCTTAAGCAGTGACCAAGCCCAACAAATAGGC-TGCATTGCTT 1938

Db 484 CCAACAAGACTTAGCTAAGCAGTGACAGACCCACCAAAAAAATAGGCTTGCAATTGCTT 543  
Qy 1939 TGAAGAG 1945  
Db 544 TGAAGAG 550  
RESULT 15  
AAS24233  
ID AAS24233 standard; cDNA; 544 BP.  
XX AC AAS24233;  
XX 07-NOV-2001 (first entry)  
DE Human ovarian PCR-subtracted cDNA library clone #414.  
XX Immunogenic protein; cancer; ovarian tumour; T-cell stimulation; ss;  
KW gene therapy; cytostatic; T-cell expansion; nucleic acid hybridisation;  
KW primer; probe.  
XX Homo sapiens.  
OS  
XX  
XX PN WO200157207-A2.  
XX  
XX PD 09-AUG-2001.  
XX PF 05-FEB-2001; 2001WO-US03733.  
XX PR 04-FEB-2000; 2000US-0180403.  
XX PR 28-MAR-2000; 2000US-0192745.  
XX  
XX PA (CORI-) CORIXA CORP.  
XX  
XX PI Algate PA, Mannion J;  
XX WPI; 2001-488879/53.  
DR  
XX New polynucleotides encoding ovarian tumour proteins, useful for  
PT treating ovarian cancer, and as probes, primers, and markers of cancer  
PT progression -  
XX  
XX Example 1; page 179; 378pp; English.

CC The invention comprises compositions used for the therapy and diagnosis  
CC of ovarian cancer. The compositions comprise one or more ovarian tumour  
CC proteins, their associated polynucleotides, or immunogenic portions of  
CC the proteins. The ovarian tumour polynucleotides and polypeptides are  
CC useful for stimulating and/or expanding T cells specific for a tumour  
CC protein. They are also useful for inhibiting the development of cancer in  
CC a patient with an ovarian tumour DNA or protein by incubating isolated  
CC T-cells allowing them to proliferate, and administering to the patient.  
CC The sequences can be used as markers for cancer, for example, to monitor  
CC ovarian cancer progression. Probes and primers are useful in nucleic acid  
CC hybridisation, in detecting the presence of complementary sequences in a  
CC given sample, for preparing mutant species and for preparing other  
CC genetic constructions. Sequences AAS23820-AAS25231 and AAS25328-AAS25549  
CC represent human ovarian tumour protein cDNA clones.

XX Sequence 544 BP; 159 A; 135 C; 130 G; 119 T; 1 other;

Query Match 18.5%; Score 526.4; DB 22; Length 544;  
Best Local Similarity 99.6%; Pred. No. 2e-86;  
Matches 538; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 1407 CCTGGCTTCTCGGGATGCTCCAGAACAAAGGACTAACGACTACTGCTTTGACTATAAC 1466  
Db 1 CCTGGCTTCTCGGGATGCTCCAGAACAAAGGACTAACGACTACTGCTTTGACTATAAC 60  
Qy 1467 CCTCCCGATGAACACAGATTGTGGACACCCAGGTCAATCTGTACCTCTGTCATGGGATG 1526  
Db 61 CCTCCCGATGAACACAGATTGTGGACACCCAGGTCAATCTGTACCTCTGTCATGGGATG 120

Qy 1527 GGCCAGAAATCAGTTTTTTCGAGTACAGTCCCGAGAAAGAAATACGCTATATACACCCACCAG 1586  
Db 121 GGCCAGAAATCAGTTTTTTCGAGTACAGTCCCGAGAAAGAAATACGCTATATACACCCACCAG 180  
Qy 1587 CCTGAGGGCTGCATTGCTGTGGAGCAGGAATGGATACCCCTTATCATGCAATCTCTGGGAA 1646  
Db 181 CCTGAGGGCTGCATTGCTGTGGAGCAGGAATGGATACCCCTTATCATGCAATCTCTGGGAA 240  
Qy 1647 GAAACTGCCCCAGAGAAATCAGAAAGTTTCATCTTGCAGGAGGATGATCTTTTATTTTCAGAA 1706  
Db 241 GAAACTGCCCCAGAGAAATCAGAAAGTTTCATCTTGCAGGAGGATGATCTTTTATTTTCAGAA 300  
Qy 1707 CAGTCCAAAGAAATGTGTCCAGGCTCGAGGAGGAGTCCAGTGACAGTTTTCGTTCCACTC 1766  
Db 301 CAGTCCAAAGAAATGTGTCCAGGCTCGAGGAGGAGTCCAGTGACAGTTTTCGTTCCACTC 360  
Qy 1767 TTACGAGACTGCACCAACTCCGATCATCAGAAATGGTTCTTCAAGAGCGCATGTTATGA 1826  
Db 361 TTACGAGACTGCACCAACTCCGATCATCAGAAATGGTTCTTCAAGAGCGCATGTTATGA 420  
Qy 1827 AGCCTCGTGTATCAAGGAGGCCCATCGAAGGAGACTGTGGAGCCGAGGACTCTGCCCAACAA 1886  
Db 421 AGCCTCGTGTATCAGAGAGCCCATCGAAGGAGACTGTGGAGCCGAGGACTCTGCCCAACAA 480  
Qy 1887 AGACTTAGCTAAGCAGTACACGAAACCCACCAAAAATACTAGGC-TGCATTTGCTTTGAAGAG 1945  
Db 481 AGACTTAGCTAAGCAGTACACGAAACCCACCAAAAATACTAGGCTTGCATTTGCTTTGAAGAG 540

Search completed: December 7, 2003, 13:57:46  
Job time : 742.064 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 7, 2003, 13:31:08 ; Search time 161.847 Seconds  
(without alignments)  
7772.402 Million cell updates/sec

Title: US-10-074-527-1  
Perfect score: 2850  
Sequence: 1 ccggctcggtaccactataa.....aaaaaaaaaaaaaaaaaaaaa 2850

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0.

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.\*

- 1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq.\*
- 2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq.\*
- 3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq.\*
- 4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq.\*
- 5: /cgn2\_6/ptodata/2/ina/PCTUS.COMB.seq.\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	496.2	17.4	1737	4	US-09-217-306B-1
2	285.6	10.0	1920	3	US-09-376-856-1
3	284	10.0	3332	3	US-09-347-488-1
4	253.8	8.9	1521	4	US-09-795-926-30
5	253.8	8.9	1812	4	US-09-795-926-42
6	253.8	8.9	3896	4	US-09-795-926-44
7	232.2	8.1	1617	2	US-08-967-508-18
8	232.2	8.1	1617	3	US-08-967-506-18
9	232.2	8.1	1617	5	PCT-US94-02552-18
10	232.2	8.1	1680	2	US-08-967-508-11
11	232.2	8.1	1680	3	US-08-967-506-11
12	232.2	8.1	1680	5	PCT-US94-02552-11
13	232.2	8.1	2294	2	US-08-967-508-10
14	232.2	8.1	2294	3	US-08-967-506-10
15	232.2	8.1	2294	5	PCT-US94-02552-10
16	222	7.8	3889	2	US-08-648-298-1
17	201.4	7.1	810	4	US-09-795-926-26
18	201.4	7.1	1101	4	US-09-795-926-38
19	198.8	7.0	1608	4	US-09-795-926-28
20	198.8	7.0	1899	4	US-09-795-926-40
21	134.8	4.7	675	4	US-09-795-926-22
22	134.8	4.7	966	4	US-09-795-926-34
23	97.6	3.4	903	4	US-08-620-312D-938
24	87.6	3.1	322	1	US-08-594-031-61
25	76.4	2.7	262	1	US-08-594-031-62
26	74.2	2.6	2416	4	US-09-620-312D-112
27	69.2	2.4	339	4	US-09-795-926-24

#### ALIGNMENTS

##### RESULT 1

US-09-217-306B-1  
; Sequence 1, Application US/09217306B  
; Patent No. 6465220

; GENERAL INFORMATION:

; APPLICANT: Hassan, Helle

; APPLICANT: Clausen, Henrik

; APPLICANT: Bennett, Eric P.

; TITLE OF INVENTION: Glycosylation Using GalNac-T4 Transferase

; FILE REFERENCE: 8850\*1

; CURRENT APPLICATION NUMBER: US/09/217,306B

; CURRENT FILING DATE: 1998-12-21

; NUMBER OF SEQ ID NOS: 25

; SOFTWARE: Patent in version 3.1

; SEQ ID NO 1

; LENGTH: 1737

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-217-306B-1

Query Match	17.4%;	Score 496.2;	DB 4;	Length 1737;
Best Local Similarity	62.3%;	Pred. No. 5.2e-96;		
Matches	796;	Conservative	0;	Mismatches 478;
			Indels	3;
			Gaps	1;
Qy	300	GCCTGGCGCGCGGGCGAGCGGTGCGGCTGCAGCTGCAGGGCGAGGAGTGGCGCTG	359	
Db	217	GCACCTTGGGAGTGGGGAAAGCCAGCAAACTCCAGCTCAACGAGGATGAATGAACGACG	276	
Qy	360	CAGGAGGAGAGCGTGGCGCTGCACACAGATTAACATCTACCTCAGCGACCCGCTCACTG	419	
Db	277	CAAGAAGAACTCAATTGAGAGATACGCCATCAATATTTACCTCAGTACAGGATTTCCCTG	336	
Qy	420	CACCGCGCGCTGCCGAGCGCTGGAACCGCGCTGCAAGAGAGAAATATGATATGAT	479	
Db	337	CATCGACACATAGAGGATAAAGATGATAGTGTAAGTCCAGAGATTCACATATAGG	396	
Qy	480	AATTTGCCAGGACATCTGTATCATAGCATTTTATAAGCAAGCTGGTCACTCTCTT	539	
Db	397	ACACTTCTTACCACCTCTGTATCATTTATTTATTAACGAGCTGGTGCAGCTTCTCTC	456	
Qy	540	CGGACAGTTTACAGTCTCTTGGAGACATCCCGGATATCTCTCGTAGAAGATGATCTT	599	
Db	457	CGTACCATTACAGTGTTTTAGAAACTTCTCTCGAGTCTTTTGTAAAGAGATCATCTG	516	
Qy	600	GTAGATGACTACAGTATAGAGAGACCTGGAAGCGCTTGGCCAAATGAGCTTTCGGA	659	
Db	517	GTGATGACTTTGAGTGACAGAGTTTATTGGAAGACACAACTTGAACATTCACAGCAAT	576	
Qy	660	CTGCCCAAGGTGCGCTGATCCGCGCAACAAGAGAGAGCGCTGTGTCGAGCCGCGCTG	719	





QY 1290 TTGGAACCTTTTGGGGATGTGACAGAGAGG-----AGCAGCTCCGGGACAAAGCTC 1340  
DB 1381 TCCTTGAGCAAGGCTGAGAAAGCCAGACTGCTGATGGAACGCTTGAGCTGCAAAAGGAGACTG 1440  
QY 1341 CAGTGTAAAGACTTTCAAGTGGTCTTCTGGAGACTGTGTATCCAGAACTGCATGTGCTGAG 1400  
DB 1441 GGTGTGGACATTTCCACTGGTCTTCTGGCTTAATGTCTACCTTGAGCTGTACCCATCTGAA 1500  
QY 1401 GACAGGCTGGCTTTCTTGGGATGCTCCAGAAACAAAGGACTTAACAGACTACTGCTTTGAC 1460  
DB 1501 CCCAGGCCAGTTTCTCTGGAAGCTCCACAACTGAGCT-----TGGGCTCTGT 1551  
QY 1461 TATAACCTCCGATGAAACACAGATTGTGGGACACAGAGTCAATCTGTACCTCTGTFCAT 1520  
DB 1552 GCAGACTGCCAGGACAGAGGAGCATCTGGGCTGTCCCATGTGTGGTCTCTCTTGCAGT 1611  
QY 1521 GGGATGGCCAGAAATCAGTTTTCGAGTACAGTCCCGAGAAAGAAATACGCTATACACC 1580  
DB 1612 GACAGCCGGACGACACAGTACCTGACACACACAGGAGGAGATTCATCTTGGCAGC 1671  
QY 1581 CACCAGCTGAGGCTGCATTGTGTGGAAGCAGGAATGGATACCTTTATCATGCTCTC 1640  
DB 1672 CCACAGC---ACCTGTGCTTGTGTGAGCAGGAGCAGTGATTCTTCAGAACTGCACG 1728  
QY 1641 TGGAGAAACTGCCCCAGAGAAATCAGAAATTCAATCTTGGAGGAGTGGATCTTTATTT 1700  
DB 1729 GAGGAAGGCTGGCCATCCACAGCAGCACTGGGACTTCCAGGAGAAATGGGATGTGTC 1788  
QY 1701 CACGAACAGTCCAAGAAATCTGCCAGGCTGCGAGGAGGAGTGCAGTGACAGTTT 1756  
DB 1789 CACATTTCTTGGGAATGCATGGAAGCTGTGTGCAAGAAACAAATAAAGATTT 1844

## RESULT 3

US-09-347-488-1

; Sequence 1, Application US/09347488  
; Patent No. 6239266  
; GENERAL INFORMATION:  
; APPLICANT: Pribill, Ingrid  
; APPLICANT: Munroe, David  
; TITLE OF INVENTION: ZAP-3 TUMOR ASSOCIATED GENES AND THEIR  
; FILE OF INVENTION: USES  
; FILE REFERENCE: GEN-2PRV  
; CURRENT APPLICATION NUMBER: US/09/347,488  
; CURRENT FILING DATE: 1999-07-02  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 3332  
; TYPE: DNA  
; ORGANISM: H. sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (368)...(2284)  
US-09-347-488-1

Query Match 10.0%; Score 284; DB 3; Length 3332;  
Best Local Similarity 53.8%; Pred. No. 7.4e-51;  
Matches 740; Conservative 0; Mismatches 600; Indels 36; Gaps 6;  
QY 399 CTCAGCGACCGCATCTCACTGCAACCGCGCTCCCGAGCGCTGGAAACCGCTGTGCATA 458  
DB 854 CTCAGTGGCCGATCCCGCTCCAGAGGGCTCTCCCGAGGTGCGGACCCCACTGTG--T 910  
QY 459 GAGAAGAAATATGATATGATTAATTTGCCAGAGACATCTGTTATCATAGATTTTAAAT 518  
DB 911 CTCAGCAGACCCCTCAGGACAGCCCTGCCACAGCCAGCGCTATCTCTGTTTCCATGAT 970  
QY 519 GAAGCTGTGTCAACTCTCTCTCGGACAGTTTACAGTGTCTCTGAGACATCCCGGATATC 578  
DB 971 GAGGCTGTGTCTCTCTCGGACTGTACAGCATCTCTCGACACAGTGTCACAGGGCC 1030

QY 579 CTCTAGAAAGAGTGTATCTTGTAGATGACTACAGTATAGAGAGCACTTGAAGGACGCC 638  
DB 1031 TTCTTGAAGGAGATCATCTCTGCTGACGACCTCAGCGAGCAAGCAACTCAAGCTCTGCT 1090  
QY 639 TTGGCCAAATGAGCTTTTCGGGACTGCCCAAGTGGCTGATCCGCGCAACAGAGAGAG 698  
DB 1091 CTGAGCGAATATGTGGCCAGGCTGGAGGGGTGAAGTTACTCAGGAGCAACAAGAGGCTG 1150  
QY 699 GGCTGTGTGAGCCCGGCTGTCTGGGGGCGCTCTGCGCGAGGGGCGCATGTTCTTGACCTTC 758  
DB 1151 AGTGCCATCAGGCGCCGATGCTGGGGGCCACAGAGCCACCGGGATGTGCTCGTCTTC 1210  
QY 759 CTGGAATCTCAGTGTGAGTGCACGAAAGGTGTGAGCGCTGAGCGCTGTGCGAGAGATCCAT 818  
DB 1211 ATGGATGCCCACTGCGAGTGCACCCAGGCTGGCTGGAGCCCTCTCTCAGCAGAATAGCT 1270  
QY 819 GAAGAGAGTCCGAGTGTGTGCCGCTGATTTGATGTGATCAGTGCAGACACCTTCGAA 878  
DB 1271 GGTGACAGAGCCGAGTGTGTTCTCCGTTGATAGATGTGATTGACTGGAAGACTTTCCAG 1330  
QY 879 TACCTGGGGAACCTCCGGGAGCCCCAGATCGGCGGTTTCGACTGGAGGCTGGTGTTCACG 938  
DB 1331 TA---TTACCCCTCAAGGACCTGCGAGCTGGGGTGTGAGCTGGAAGCTGGATTTCCAC 1387  
QY 939 TGGCACACAGTTCTCTGAGAGGAGAGATACGGATGCAATCCCGCTGCGATGTCATCAGG 998  
DB 1388 TGGGAACCTTTTGCAGAGCATGTGAGGAAGGCCCTCCAGTCCGCCAATAAGCCCATCAGG 1447  
QY 999 TCTCCAAACATGGCTGGTGGCTGTTTCTGTGAGTAAGAAATATTTTGAATATCTGGGG 1058  
DB 1448 AGCCCTGTGTGCTCCGAGAGGTGGTGGCCATGGACAGACATTAATTTCCAAAACACTGGA 1507  
QY 1059 TCTTATGATACAGGAATGGAAGTTTGGGGAGGAGAAACCTCGAATTTTCTCTTAGGATC 1118  
DB 1508 GCGTATGACTCTTATGTGCTGCGAGGTGGTGAAGAACTCGAATCTCTTTCAAGGCC 1567  
QY 1119 TGGCAGTGTGGGGTCTGGAACACACACCCATGTTCCCATGTTGGCCATGTTTCCCC 1178  
DB 1568 TGGCTCTGTGGTGGCTCTGTTGAAATCTTCCCTGCTCTCGGCTAGGACATCTACCAA 1627  
QY 1179 AGCAGCTCCCTACTCTCCGCAACAGGCTCTGGCC-----ACAGTGTTCGTGCA 1229  
DB 1628 AATCAGGATTCCTCATTTCCCGCTCGACAGGAGGCCACCTGAGGAACAGGGTTCGCATT 1687  
QY 1230 GCTGAAGTATGGATGAATTTAAAGAGCTCTACTACCTCGCAACCCCGCTGCCCGC 1289  
DB 1688 GCTGAGACCTGCTGGGGTCAATCAAGAAACCTTTACAGCATAGCCAGAGGCCCTTC 1747  
QY 1290 TTGGAACCTTTTGGGGATGTGACAGAGAGGA-----AGCAGTCCGGGACAAAGCTC 1340  
DB 1748 TCCTTGAGCAAGGCTGAGAAAGCCAGACTGCTGGAACGCTTGCAGCTGCAAGGAGACTG 1807  
QY 1341 CAGTGTAAAGACTTCAAGTGGTCTTGGAGACTGTGTATCCAGAACTGCATGTGCTGAG 1400  
DB 1808 GGTGTGCGGACATTCACACTGGTCTTGGGCTTAATGTCTACCTGAGCTGTACCCATCTGAA 1867  
QY 1401 GACAGGCTGGCTTCTTTCGGGATGCTCCAGAACAAAGGACTTAACAGACTACTCTGTTGAC 1460  
DB 1868 CCAGGCCAGATTTCTCTGAAAGCTCCACACACTGGACT-----TGGGCTCTGT 1918  
QY 1461 TATAACCTCCCGATGAAACCAAGATTGTGGGACACAGGCTCAATCTGTACCTCTGTFCAT 1520  
DB 1919 GCAGACTGCCAGCAGAGGGGACATCTCTGGGCTGTCCTCATGTTGGCTCTCTTGCAGT 1978  
QY 1521 GGGATGGGCGAGAAATCAGTTTTTCGAGTACACGTCCAGAAAGAAATACGCTATAACACC 1580  
DB 1979 GACAGCCGCGACCAACAGTACCTGCGAGCACACCCAGCAGGAAGGAGATTCACTTTGGCAGC 2038  
QY 1581 CACAGCCTTGAGGGCTGCAATGCTGTGGAAGCAGGAATGATACCTCTATCATGCTATC 1640  
DB 2039 CCACAGC---ACCTGTGCTTGTCTGAGGAGGAGCTGATTTCTTCAGAACTGCACG 2095  
QY 1641 TGGGAAGAAACTGCCCCCAGAGAAATCAGAAAGTTTCATCTTGCAGGAGGATGATCTTTATTT 1700

Db 2096 GAGGAGGCGCTGCCATCCACAGCAGCACTGGGACTTCACAGAGATGGGATGTC 2155  
Qy 1701 CACGAACAGTCCAGAAATGTGTCAGGCTGCGAGGAGAGATCGAGTGCAGTTT 1756  
Db 2156 CACATTCTTTCTGGAAATGCATGGAAGCTGTGTGCAAGAAACAAATAAGATT 2211

RESULT 4

US-09-795-926-30  
; Sequence 30, Application US/09795926  
; Patent No. 6555669  
; GENERAL INFORMATION:  
; APPLICANT: Donoho, Gregory  
; APPLICANT: Hilbun, Erin  
; APPLICANT: Turner, C. Alexander Jr.  
; APPLICANT: Friedrich, Glenn  
; APPLICANT: Abuin, Alejandro  
; APPLICANT: Zambrowicz, Brian  
; APPLICANT: Sands, Arthur T.  
; APPLICANT: Walke, D. Wade  
; APPLICANT: Wilganowski, Nathaniel L.  
; APPLICANT: Hu, Yi  
; APPLICANT: Kieke, James Alvin  
; APPLICANT: Potter, David George  
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
; FILE REFERENCE: POLYNUCLEOTIDES ENCODING THE SAME  
; CURRENT FILING DATE: 2001-02-28  
; PRIOR FILING DATE: 2001-02-28  
; PRIOR FILING DATE: 2000-02-29  
; PRIOR FILING DATE: 2000-03-02  
; PRIOR FILING DATE: 2000-03-02  
; PRIOR FILING DATE: 2000-03-24  
; NUMBER OF SEQ ID NOS: 47  
; SOFTWARE: PASTSEQ for Windows Version 4.0  
; SEQ ID NO 30  
; LENGTH: 1521  
; TYPE: DNA  
; ORGANISM: homo sapiens  
US-09-795-926-30

Query Match 8.9%; Score 253.8; DB 4; Length 1521;  
Best Local Similarity 55.5%; Pred. No. 1.3e-44;  
Matches 561; Conservative 0; Mismatches 432; Indels 18; Gaps 3;  
Qy 388 TTAACATCTACTCAGCGACCGCATCTCACTGCACCGCGCTGCCCGAGCGCTGGAACC 447  
Db 47 TTAACATCTAGTCAGTGAATTAATCTCTTGAATCGCTCTCTCCAGATATCCGGCACC 106  
Qy 448 CGCTGTGCAAGAGAAATATGATTATGATAATTTGCCAGGACATCTGTTATCATAG 507  
Db 107 CAACTGCAACAGCAGCGCTACCT--GGAGACACTTCCCAACACAGCATCATATCC 163  
Qy 508 CATTTTATATGAAGCCTGTCATCTCTCTCGGACAGTTTACAGTGTCTTGAGACAT 567  
Db 164 CTTTCCACACAGGCGGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 223  
Qy 568 CCCCAGATATCTCTAGAGAGATGATCTCTGTGATGATGATGATGATGATGATGATGATG 627  
Db 224 CGCTCCAGAGCTGTGTCAGAGATTTGATGTCGAGACTTCAGTATCGAGGACACC 283  
Qy 628 TGAAGAGCGCTTGGCCAAATGAGCTTTTGGGACTGCCCAAGGTGCGCTGATCCGCCCA 687  
Db 284 TGAAGAGCCTTTGAAGACTACATGCGCTTTTCCCGAGTGTGAGGATCTTCCGAACCA 343  
Qy 688 ACAAGAGAGGGGCTGTGTCAGCGCGCTGTGTCGGGCGCTCTGCGCGAGGGCGATG 747  
Db 344 AGAAACGGGAGGGCTGTGTAAGAGACCCCGAATGCTGGGGCGCTCAGTGCAACTGGGATG 403  
Qy 748 TTCTGACCTTCTGCACTGTCACTGTGAGTGCCACGAGGGGTGGCTGGAGCGCGCTGCTGC 807

Db 404 TCATCACTTCTTGGATTCACTGTGAAGCAATGTCAACTGGCTTCCCGCTTGGTTG 463  
Qy 808 AGAGGATCCATGAAGAGGAGTGGCAGTGGTGGCCGGTGTGATGATGATGATGATGATG 867  
Db 464 ACGCATCTCTCGGAACCGCAAGACCAATTTGTGTGCGCGATGATGATGATGATGATG 523  
Qy 868 ACACCTTTCGAATACCTCGGGGAACTCCGGGAGGAGCCAGATCGGCGGTTTCGACTGGAGGC 927  
Db 524 ACGATTTTCGGTACGAGACACAGGAGGGATGCCATCGGGGAGGCTTTGACTGGGAGA 583  
Qy 928 TGGTGTTCACGTGGCACACAGTTTCTCGAGAGGAGAGATACGATGCAATTCGCCGTCG 987  
Db 584 TGTACTACAAGCGGATCCCGATCCCT-----CCAGAACTGCAGAAAGCTGACCCAGCG 637  
Qy 988 ATGTATCATAGTCTCCAAACATGGCTGTGGGCTGTGCTGTGATGATGATGATGATGATG 1047  
Db 638 ACCATTTGAGTCTCCCGTGTGATGGCGGTGGACTGTTCCGCGTGGATCGGAAGTGGTTCT 697  
Qy 1048 AATATCTGGGCTCTTATGATACAGGAATGGAAGTTTGGGAGGAGAGAAACCTCGAATTTT 1107  
Db 698 GGAATCTGGCGGTATGACCCAGGCTTGGAGATCTGGGAGGGGAGCAGTATGAAATCT 757  
Qy 1108 CTTTAGGATCTGGCAGTGTGGTGTCTTGGAAACACACCATGTTTCCCATGTTGGCC 1167  
Db 758 CTTCAAGGTGTGATGTGTGGGGCGCATGAGGACATCCCTCTGCTCCAGGTTGGGCC 817  
Qy 1168 ATGTTTTCCCAAGCAAGCTCCCTAC-----TCCCGCAACAGGCTCTGCCCAACA 1218  
Db 818 ATATCTACAGGAATGATGTGCCCTCAAGGTCCCGCGGAGTCAAGCTGGCGCGGAAACC 877  
Qy 1219 GTCTTCGTGAGCTGAAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 1278  
Db 878 TTAAGCGGTGGCGGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATG 937  
Qy 1279 CCGTGTCCCGCTTGGAACTTTTGGGGATGTGACAGAGAGAGAGCAGCTCCGGGACAGC 1338  
Db 938 CTGAATAGCGCCACCTCTCCGCTGGGATGTGTCAGTCCAGAAAGCTCCGAGCTCCC 997  
Qy 1339 TCCAGTGTAAAGACTTCAAGTGTCTTGGAGACTGTGATCCAGACTGC 1389  
Db 998 TTAACCTCAAGAGTTCCTCAAGTGTGTTATGACGAAGATAGCTGGGACCTGC 1048

RESULT 5

US-09-795-926-42  
; Sequence 42, Application US/09795926  
; Patent No. 6555669  
; GENERAL INFORMATION:  
; APPLICANT: Donoho, Gregory  
; APPLICANT: Hilbun, Erin  
; APPLICANT: Turner, C. Alexander Jr.  
; APPLICANT: Friedrich, Glenn  
; APPLICANT: Abuin, Alejandro  
; APPLICANT: Zambrowicz, Brian  
; APPLICANT: Sands, Arthur T.  
; APPLICANT: Walke, D. Wade  
; APPLICANT: Wilganowski, Nathaniel L.  
; APPLICANT: Hu, Yi  
; APPLICANT: Kieke, James Alvin  
; APPLICANT: Potter, David George  
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
; FILE REFERENCE: POLYNUCLEOTIDES ENCODING THE SAME  
; CURRENT FILING DATE: 2001-02-28  
; PRIOR FILING DATE: 2001-02-28  
; PRIOR FILING DATE: 2000-02-29  
; PRIOR FILING DATE: 2000-03-02  
; PRIOR FILING DATE: 2000-03-24  
; NUMBER OF SEQ ID NOS: 47

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 42

; LENGTH: 1812

; TYPE: DNA

; ORGANISM: homo sapiens

US-09-795-926-42

Query Match 8.9%; Score 253.8; DB 4; Length 1812;

Best Local Similarity 55.5%; Pred. No. 1.4e-44;

Mismatches 561; Conservative 0; Mismatches 432; Indels 18; Gaps 3;

```
Qy 388 TTAACATCTACCTCAGCGACCGCATCTCACTGCAACCGCGCTGCGCGAGCGCTGGAACC 447
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 338 TTAACATCTACCTCAGCGACCGCATCTCACTGCAACCGCGCTGCGCGAGCGCTGGAACC 397
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 448 CGCTGTGCAAGAGAGAAATATGATTATGATTAATTTGCCAGGACATCTGTTATCATAG 507
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 398 CAAACTGCAACAGCAAGCGCTACCT---GGAGACATTTCCCAACACAGCATCATATCC 454
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 508 CATTTTATTAATGAAGCTGTGTAACCTCTCTTGGGACAGTTTACAGTGTCTTGAGACAT 567
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 455 CCTTCCAAACAGAGGGTGGTCTCTCTCTCCGACCGCTCCACAGTGTGCTCAATCGCT 514
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 568 CCCGGATATCTGCTAGAGAGTGTATCTTGTAGATGACTACAGTATAGAGAGCACC 627
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 515 CGCTCCAGAGCTGTGCGCGAGATTGTAATGCTGCGAGACTTCAGTATCGAGAGCACC 574
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 628 TGAAGGAGCGCTTTGGCCAATGAGCTTTTCGGGACTGCCCCAAGGTGCGCTGATCCGCGCCA 687
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 575 TGAAGAGCCTCTTGAAGACTACATGCGCCCTTTTCCCGAGTGTGAGGATTTTCGACCA 634
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 688 ACAAGAGAGAGGCTGTGTCGAGCCGCGTGTGCGGGCGTCTGCGGAGAGGGCGATG 747
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 635 AGAAACGGGAGGCTGTAAGGACCGCAATGCTGCGGGGCTCAGTGGCAACTGGGGATG 694
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 748 TTCTGACCTTCCTGGAAGTGTCACTGTGAGTGCACAGAGGGTGGCTGGAGCCGCTGTCG 807
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 695 TCATCAATTTCTGGATTACACTGTGAAGCAATGTCAATGGCTTTCCCGCCCTTGGTTG 754
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 808 AGAGGATCCATGAAGAGGAGTGGCGAGTGGTGTGCGCGGTGATTGATGTGATCGACTGGA 867
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 755 ACCGATTTGCTCGGAACCGCAAGACCAATTTGTGCGCGCGATGATTGATTAATGACATG 814
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 868 ACACCTTCGAATACCTGGGAACTCCGGGAGCGCCAGATCGCGGTTTCAGCTGGAGGC 927
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 815 ACCGATTTCCGTACGAGACACAGCAGCGGATGCCATGCGGGAGCCTTTGACTGGGAGA 874
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 928 TGGTGTTCAGTGGCAGACAGTTCCTGAGAGGAGAGGATACGGATCAATCCCGCTCG 987
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 875 TGTACTACAAGCGGATCCCGATCCCT-----CCAGAACTGCAGAAAGCTGACCCAGCG 928
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 988 ATGTCTACAGTCTCCAAACAATGGCTGGTGGGCTGTTTGTGTGAGTAAGAAATATTTTG 1047
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 929 ACCAATTTGAGTCTCCGCTGATGGCGGTGAGATGTTGCGCGGTGATCGGAAGTGTCT 988
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1048 AATATCTGGGCTTTATGATACAGGAATGGAAGTTTGGGAGGAGAAACCTCGAATTTT 1107
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 989 GGGAACTCGCGGGGTATGACCCAGGCTTGAGATCTCGGGAGGGAGCAGTATGAATCT 1048
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1108 CTTTAGGATCTGGCAGTGTGGGGTCTTGGAAACACACCCATGTTCCCATGTTGGCC 1167
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1049 CTTCAAGGTGTGATGTGGGGGCGCATGGAGGACATCCCGCTGCTCCAGGTTGGGCC 1108
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1168 ATGTTTTCCTCAAGAGCTCCCTAC-----TCCGCAACAAGGCTCTGGCAACA 1218
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1109 ATATCTACAGAGATATGTCCTTACAGGTCCCGCGCGAGTCAAGCTGCGCGCGGAACC 1168
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1219 GTGTTCTGTCAGTGAAGTATGATGATGAATTTAAAGAGCTCTACTACCATCGCAACC 1278
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1169 TTAAGCGGGTGGCGAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATG 1228
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1279 CCGTGGCCGCTTTGGGAACCTTTTGGGGATGTGACAGAGAGAGAGAGCTCCGGGACAA 1338
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
Db 1229 CTGAATACCGCCACCTCTCCGCTGGGGATGTCGAGTCCAGAAAAAGCTCCGAGCTCCC 1288
Qy 1339 TCCAGTGTAAAGACTTCAAGTGGTCTTCTTGAGAGACTGTATCCAGAACTGC 1389
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1289 TTAAGTCAAGAGTTCACAGTGGTATTATGACGAAGATAGCTGGGACCTGC 1339
```

#### RESULT 6

US-09-795-926-44

; Sequence 44, Application US/09795926

; Patent No. 6555669

; GENERAL INFORMATION:

; APPLICANT: Donoho, Gregory

; APPLICANT: Hilbun, Erin

; APPLICANT: Turner, C. Alexander Jr.

; APPLICANT: Friedrich, Glenn

; APPLICANT: Abuin, Alejandro

; APPLICANT: Zambrowicz, Brian

; APPLICANT: Sands, Arthur T.

; APPLICANT: Walke, D. Wade

; APPLICANT: Wilganowski, Nathaniel L.

; APPLICANT: Hu, Yi

; APPLICANT: Kieke, James Alvin

; APPLICANT: Potter, David George

; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND

; FILE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME

; FILE REFERENCE: LEX-0144-USA

; CURRENT APPLICATION NUMBER: US/09795,926

; PRIOR FILING DATE: 2001-02-28

; PRIOR APPLICATION NUMBER: US 60/185,920

; PRIOR FILING DATE: 2000-02-29

; PRIOR APPLICATION NUMBER: US 60/186,558

; PRIOR FILING DATE: 2000-03-02

; PRIOR APPLICATION NUMBER: US 60/191,849

; PRIOR FILING DATE: 2000-03-24

; NUMBER OF SEQ ID NOS: 47

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 44

; LENGTH: 3896

; TYPE: DNA

; ORGANISM: homo sapiens

US-09-795-926-44

Query Match 8.9%; Score 253.8; DB 4; Length 3896;

Best Local Similarity 55.5%; Pred. No. 2e-44;

Mismatches 561; Conservative 0; Mismatches 432; Indels 18; Gaps 3;

```
Qy 388 TTAACATCTACCTCAGCGACCGCATCTCACTGCAACCGCGCTGCGCGAGCGCTGGAACC 447
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 347 TTAACATCTACCTCAGCGACCGCATCTCACTGCAACCGCGCTGCGCGAGCGCTGGAACC 406
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 448 CGCTGTGCAAGAGAGAAATATGATTATGATTAATTTGCCAGGACATCTGTTATCATAG 507
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 407 CAAACTGCAACAGCAGCGCTACCT---GGAGACATTTCCCAACACAGCATCATATCC 463
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 508 CATTTTATTAATGAAGCTGTGTAACCTCTCTCGGACAGTTTACAGTGTCTCTTGAGACAT 567
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 464 CTTTCCCAACAGAGGCTGGTCTCTCTCTCGCACCGTCCACAGTGTCTCAATCGCT 523
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 568 CCCCAGATATCTGTCTAGAGAAAGTGTATCTTGTAGATGATACAGTGTAGAGAGCACC 627
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 524 CGCTCCAGAGCTGTGCGCGAGATTGTAATGTAATGTAATGTAATGTAATGTAATGTAATG 583
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 628 TGAAGAGCGCTTGGCCAATGAGCTTTGGGACATGCGGACCTGCGGAGTGTGATCGCGCCA 687
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 584 TGAAGAGCGCTTGAAGACTACATGGCCCTTTTCCCGAGTGTGAGGATTTCTTCGAACCA 643
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 688 ACAAGAGAGAGGCGCTGGTGGCGAGCCGCGTGTCTGGGGCGCTCTGCGGCGAGGGCGATG 747
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 644 AGAAACGGGAGAGGCTGATAAGGACCGCAATGCTGGGGGCTCAGTGGCAACTGGGGATG 703
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 748 TTCTGACCTTCCTGGAAGTGTCACTGTGAGTGCACAGAGAGAGAGCTCCGGGACAAAGC 807
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Db 704 TCATCATCTTTGGATTACACTGTGAAGCAATGTCAACTGGCTTCCCGCTTGGCTG 763  
Qy 808 AGAGGATCCATGAAGAGAGTGGCAGTGTGTGCGCGGTGATTGATGTGATGATGATGATG 867  
Db 764 ACCGATTTGTCGGAACCGCAAGACCATTTGTGCGCGGATGATGATGATGATGATGATG 823  
Qy 868 ACACCTTTCGAATACCTCGGGAACTCCGGGAGCCCGAGATCGCGGTTCGACTGGAGGC 927  
Db 824 ACGNCTTCGTCAGACACACGCGAGGAGTCCATGCGGGAGCCCTTGTGATGCGGAGA 883  
Qy 928 TGGTGTTCAGTGGCACAACAGTTCTCGAGAGGAGAGATACCGATGCAATCCCGCTGCG 987  
Db 884 TGTACTACAAGCGGATCCCGATCCCT-----CCAGAACTGCAGAAAGCTGACCCACGC 937  
Qy 988 ATGTCATCAGGTCTCCAAACATGGCTGTGGGCTGTTTGTGCTGAGTAAGAAATATTTG 1047  
Db 938 ACCAATTTGAGTCTCCCGTGTATGGCGCGTGGACTGTTTCGCGGTGGATCGGAAGTGTCT 997  
Qy 1048 AATATCTGGGTCTTATGATACAGGAATGGAAGTTTGGGAGGAGAGAAACCTCGAATTTT 1107  
Db 998 GGGAACTCGCGGGGTATGACCCAGGCTTGAGATCTCGGAGGAGGAGCAGTATGAATCT 1057  
Qy 1108 CTTTAGGATCTGCGAGTGTGGGTCTCGGAAACACACCCATGTTCCCATGTTGGCC 1167  
Db 1058 CTTTCAAGGTGTGATGTGGGGGCGCATGGAGGACATCCCTCTGCTCCAGGTGGGC 1117  
Qy 1168 ATGTTTTCGCAAGCAAGCTCCCTAC-----TCCGCAACAAAGCTCTGCGCAACA 1218  
Db 1118 ATATCTACAGGAATGATGCGCTTACAAAGTCCCGGCGGAGTCAGCTCGCGCCGGAACC 1177  
Qy 1219 GTGTTCTGCGAGCTGAAGTATGAGTATGATGAATTTAAAGAGCTCTACTACCATCGCAACC 1278  
Db 1178 TTAAGCGGTGGCCGAAAGTGTGGATGATGATGATGATGATGATGATGATGATGATGATG 1237  
Qy 1279 CCGTGTCCGCTTGGAACTTTTGGGATGTGACAGAGAGAAAGCAGCTCGGAGCAAGC 1338  
Db 1238 CTGAATACCGCCACCTCTCCGCTGGGATGTGCGAGTCCAGAAAGCTCGCGAGCTCC 1297  
Qy 1339 TCCAGTGAAGACTTCAAGTGGTCTTGGAGACTGTGTATCCAGAACTGC 1389  
Db 1298 TTAAGTGAAGTTCGAAGTGGTTTATGACGAAGATAGCTTGGGAGCTTC 1348

## RESULT 7

US-08-967-508-18  
; Sequence 18, Application US/08967508  
; Patent No. 5910570  
; GENERAL INFORMATION:  
; APPLICANT: The Upjohn Company  
; APPLICANT: FOR U.S. PURPOSES ONLY: Elhammer, Ake P. and Homa, Fred L.  
; TITLE OF INVENTION: A Cloned DNA Encoding a UDP-galNac:  
; TITLE OF INVENTION: Polypeptide, N-Acetylglactosaminyltransferase  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pharmacia & Upjohn Company, Intellectual  
; ADDRESSEE: Property Legal Services  
; STREET: 301 Henrietta Street  
; CITY: Kalamazoo  
; STATE: Michigan  
; COUNTRY: USA  
; ZIP: 49001  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/967,508  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/602,830  
; FILING DATE: 13 No. 5910570ember 1995

; ATTORNEY/AGENT INFORMATION:  
; NAME: Darnley Jr., James D.  
; REGISTRATION NUMBER: 33,673  
; REFERENCE/DOCKET NUMBER: 4755.P CP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 616-833-2210  
; TELEFAX: 616-833-8897  
; TELEX: 224401  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1617 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; US-08-967-508-18

Query Match 8.1%; Score 232.2; DB 2; Length 1617;  
Best Local Similarity 54.8%; Pred. No. 5.2e-40;  
Matches 612; Conservative 0; Mismatches 463; Indels 42; Gaps 6;

Qy 472 ATTATGATTAATTTGCCAGGACATCTGTTATCATAGCATTTTATATGAGCCTGGTCAA 531  
Db 269 ATCCAGATAACCTTCTTACAACCAAGTGTGGTGAATTTTCCCAATAGGCTTTGGAGCA 328  
Qy 532 CTCTCTCTCGAGCAGTTTACAGTGTCTTTCAGACATCCCGGATATCTCTCTAGAAAG 591  
Db 329 CACTTCTCGAACTGTCCATAGCTCATTAATCGCTCACCAGGCACATGCTAGAGAAA 388  
Qy 592 TGATCTCTGTAGTACTACAGTGTAGAGACACCTGAAGAGCGCTTGGCCAAATGAGC 651  
Db 389 TTGTTCTAGTAGATGATGCCAGTGAAGAGACTTTTAAAAAGACCTCTAGAGAGTTACG 448  
Qy 652 TTTTCGGGACT---GCCAAGTTCGCTGATCCGCGCCACACAGAGAGAGCGCTGGTGC 708  
Db 449 TGAATAAATTTAAAGTACCCGTTTCACTGATTCGAATGGAGCAGCGTTCTGGATGATCA 508  
Qy 709 GACCCCGCTGCTGGCGGCTCTCGCGAGGCGGATGTTCTGACCTTCTCGACTGTG 768  
Db 509 GAGCTAGTTAAAGTGTGCTGTCTTAAGGCCAAGTATGATCACTTTTATAGACGCG 568  
Qy 769 ACTGTGAGTCCACGAAGGTGCTGGAGCGCTGTGTCAGAGAGATCCATGAAGAGAGT 828  
Db 569 ACTGTGAGTGCACAGTGGGTGGCTGGAGCTCTCTTAGCCAGGATCAAAACATGACAGGA 628  
Qy 829 CGGAGTGTGTCGCCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 888  
Db 629 AGACAGTGTCTGTCCCATCATAGATGTGATCATGATGATGATGATGATGATGATGATG 688  
Qy 889 ACTCCGGGAGCCCGAGATCGCGGCTTTCGACTGGAGGCTGGTGTTCACGTGGCAGACAG 948  
Db 689 GTTCTGACATGACCTA---TGGCGGTTTCACTTGAAGCTCAACTTCGCTGGTATCTG 745  
Qy 949 TTCTCTGAGAGGGAGAGG---ATACGAGTGAATCCCCCGTCGATGTCATCAGTCTCAA 1005  
Db 746 TTCCCAAGAGAGAAATGGACAGAGAAAGGTGATCGGACTTCTCTGTGAGAAACCTA 805  
Qy 1006 CAATGGCTGTGGCTGTTTGTGCTGATGATGATGATGATGATGATGATGATGATGATG 1065  
Db 806 CAATGGCAGGAGGCTTTTTCATATAGACAGAGATTTACTTTTCAGGAAATTTGGAACATATG 865  
Qy 1066 ATACAGGAATGGAAGTTTGGGAGGAGAAACCTCGAATTTTCTTTAGGATCTGGAGT 1125  
Db 866 ATGCTGGAATGGATATTTGGGAGGAGAAACCTTAGAAATTTCTTTAGGATTTGGAGT 925  
Qy 1126 GTGTGGGCTTCTGGAAACACACCCATGTTTCCCATGTTGGCCATGTTTCCCCAAGCAAG 1185  
Db 926 GTGGAGGAACTTTGGAGATTTGTTACTTGTCTCACATGTTGGACATGTTGTTTCGGAAGCTA 985  
Qy 1186 CTCCCTACTCCGCAACAAAGCTCTGGCCAAAGTGT-----TCGTGCGAG 1230  
Db 986 CACCCTACAGTTTCCAGGAGGACCGGGGAGATTTATCAATAAAAAATAACAGACGACTTG 1045

QY 1231 CTGAGTATGGATGGATGAATTTAAAGAGCTCTACTACCATCGCAACCCCGTCCCGCT 1290  
DB 1046 CAGAAGTATGGATGGATGAATTTAAAGAGCTCTACTACCATCGCAACCCCGTCCCGCT 1105  
QY 1291 TGAACCTTTTGGGAGTGTGACAGAGAGGAGCAGCTCCGGGACCAAGCTCCAGGTAAAG 1350  
DB 1106 AGGTAGATTAGGATATATCATCAAGACTTGTCTAAGGCACAACTCCAATGCAGAC 1165  
QY 1351 ACTTCAAGTGGTCTTGGAGACTGTGTATCCAGAACTGCGATGTCCTGAGGACAGGCGTG 1410  
DB 1166 CATCTCTGGTACTAGAGAAATTTTATCTGATTCTCAGATTCTCTCGTCACTATTCT 1225  
QY 1411 GCTTCTTGGGATGCTCCAGAACCAAGACTTAACAGACTACTGTTTGCATTAACCTC 1470  
DB 1226 CTTTGGGAGATAC-----GAAATGTGAAACAAATCAGTGTCTAGATAACATGGCTA 1279  
QY 1471 CCGATGAAACCAAGATTGTGGGACACCAGGTCACTTGTACCTCTGTGATGGGATGGCC 1530  
DB 1280 GAAAGAGATGAAAGTTGGA-----ATTTTAACTGTCTATGGTATGGAG 1327  
QY 1531 AGAATCAGTTTTCGAGTACACGTCCAGAAAGAAAT 1567  
DB 1328 GTAATCAGGTTTCTCTTACACTGCCCAACAAAGAAAT 1364

## RESULT 8

US-08-967-506-18  
; Sequence 18, Application US/08967506  
; Patent No. 6096512  
; GENERAL INFORMATION:  
; APPLICANT: The Upjohn Company  
; APPLICANT: FOR U.S. PURPOSES ONLY: Elhammer, Ake P. and Homa, Fred L.  
; TITLE OF INVENTION: A Cloned DNA Encoding a UDP-GalNAC:  
; TITLE OF INVENTION: Polypeptide, N-Acetylgalactosaminyltransferase  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pharmacia & Upjohn Company, Intellectual  
; ADDRESSEE: Property Legal Services  
; STREET: 301 Henrietta Street  
; CITY: Kalamazoo  
; STATE: Michigan  
; COUNTRY: USA  
; ZIP: 49001  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/967,506  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/602,830  
; FILING DATE: 13 No. 6096512ember 1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Darnley Jr., James D.  
; REGISTRATION NUMBER: 33,673  
; REFERENCE/DOCKET NUMBER: 4755.P CP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 616-833-2210  
; TELEFAX: 616-833-8897  
; TELEX: 224401  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1617 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: Double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cdna  
US-08-967-506-18

Query Match

8.1%; Score 232.2; DB 3; Length 1617;

Best Local Similarity 54.8%; Pred. No. 5.2e-40;  
Matches 612; Conservative 0; Mismatches 463; Indels 42; Gaps 6;  
QY 472 ATTATGATTAATTTGCCAGGACATCTGTTATCATAGCATTTTATAATGAAGCCCTGGTCAA 531  
DB 269 ATCCAGATACTTCTCTACCAACAGTGTGTGATTTTTCACAAATGAGGCTTGGAGCA 328  
QY 532 CTCTCTCGGACAGTTTACAGTGTCTTGGAGATATCCCGGATATCTCTAGAGAAG 591  
DB 329 CACTTCTCGAACTGTCCATAGCGTCATTAATCGCTCACCAAGSCACATGCTAGAAGAA 388  
QY 592 TGATCTTGTAGATGATCAGTATGAGAGACACCTGAAGGAGCGCTTGGCCATAGC 651  
DB 389 TTGTTCTAGTAGATGATGCGAGTGAAGAGACTTTTAAAGAGACTCTTAGAGAGTTACG 448  
QY 652 TTTTCGGGACT---GCCCAAGGTGCGCTGATCGCGCAACAAGAGAGAGCGCTGTGTC 708  
DB 449 TGAAGAAATTAAGATACCGTTTCACTGATTCGATGAGAGAGCGTCTGGATTGATCA 508  
QY 709 GAGCCCGGCTGTGGGGGCGTCTGCGGCGAGGGCGATGTTCTGACCTTCTGAGCTGTC 768  
DB 509 GAGTAGGTTAAAGGTGCTGTGTTAAAGGCCAAGTATCACCTTTTACGCGC 568  
QY 769 ACTGTGAGTCCAGAAAGGTGCTGAGCGCTGTCGAGAGATCCATGAAGAGAGT 828  
DB 569 ACTGTGAGTGCACAGTGGGGTGGAGCTCTTAGCCAGGATCAAAACATGACAGGA 628  
QY 829 CGGAGTGGTGTCCCGGTGATTGATGATCGACTGGAACACTTCGAATACCTGGGA 888  
DB 629 AGACAGTGGTCTGCTCCATCATAGATGATCATGATGACATTTTCGAGTACATGGCAG 688  
QY 889 ACTCCGGGAGCCCGGAGATCGCGGTTTCACTGAGAGGTGGTGTTCACGTGGCACACAG 948  
DB 689 GTTCTGACATGACCTA---TGGCGGGTTCACTTGAAGCTCAACTTTTCGCTGGTATCTG 745  
QY 949 TTCTGAGAGGAGAGG---ATACGATGCAATCCCGCTGATGTCATCAGTGTCTCAA 1005  
DB 746 TTCCCAAGAGAGAAATGGACAGAAAGGTGATCGGACTCTTCTGTGAGAAACACTA 805  
QY 1006 CAATGGCTGGTGGCTGTTTGTGAGTGAAGAAATATTTTGAATATCTCGGGTCTTATG 1065  
DB 806 CAATGGCAGGAGGCTTTTCAATAGACAGAGATTACTTTTCAGGAAATTTGGAACATATG 865  
QY 1066 ATACAGGAATGGAAGTTTGGGAGGAGAAACCTCGAAATTTTCTTTAGGATCTGGCAGT 1125  
DB 866 ATCTGGAATGGATATTTGGGAGGAGAGAAACCTAGAAATTTCTTTAGGATTTGGCAGT 925  
QY 1126 GTGCTGGGTTCTGGAAACACACCCATGTTCCCATGTTGGCCATGTTTCCCAAGCAAG 1185  
DB 926 GTGGAGGAACTTTGGAGATTGTTTACTTGTCTCATGTTGGACATGTTTTCGGAAGACTA 985  
QY 1186 CTCCCTACTCCCGCAACAGGCTCTGGCCACAGTGT-----TCGTGCAG 1230  
DB 986 CACCTTACAGTTTCCAGGAGGCGACGGGCGAGATTATCAATAAAATAACAGACAGCTTG 1045  
QY 1231 CTGAAGTATGGATGATGAATTTAAAGAGCTCTACTACCATCGCAACCCCGTCCCGCT 1290  
DB 1046 CAGAAGTATGGATGATGAATTTCAAGAAATTTCTTATATAATTTCTCCAGGTGTACAA 1105  
QY 1291 TGAACCTTTTGGGATGTGACAGAGAGAGCAGCTCCGGGACAAAGCTCAGTGTAAAG 1350  
DB 1106 AGGTAGATTATGGAGATATATCATCAAGACTTGGTCTAAGGCACAACTCCAATGCAGAC 1165  
QY 1351 ACTTCAAGTGGTCTTGGAGACTGTGTATCCAGAACTGCGATGTCCTGAGGACAGGCGTG 1410  
DB 1166 CATCTCTGGTACCTAGAGAAATTTTATCTGATTCTCAGATTCTCTCGTCACTATTCT 1225  
QY 1411 GCTTCTTGGGATGCTCCAGAACCAAGGACTTAACAGACTACTGTTTGCATTAACCTC 1470  
DB 1226 CTTTGGGAGAGATAC-----GAAATGTGAAACAAATCAGTGTCTAGATAACATGGCTA 1279  
QY 1471 CCGATGAAACCAAGATTGTGGGACCAAGGTCACTTGTACCTCTGTGATGGGATGGCC 1530

```

Db      1280  GAAAGAGAAATGAAAGTTGGA-----ATTTTAACTGTCATGGTATGGGAG 132
QY      1531  AGAATCAGTTTTTCGAGTACACAGCTCCAGAGAAAGAAAT 1567
Db      1328  GTAATCAGGTTTCTTCTTACACTGCCCAACAAGAAAT 1364

RESULT 9
PCT-US94-02552-18
; Sequence 18, Application PC/TUS9402552
; GENERAL INFORMATION:
; APPLICANT: Elhammer, Ake P.
; APPLICANT: Homa, Fred L.
; TITLE OF INVENTION: A Cloned DNA Encoding a UDP-GalNac:
; TITLE OF INVENTION: Polypeptide, N-Acetylgalactosaminyltransferase
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Upjohn Company, Corp. Intellectual
; ADDRESSEE: Property Law
; STREET: 301 Henrietta Street
; CITY: Kalamazoo
; STATE: Michigan
; COUNTRY: USA
; ZIP: 49001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/02552
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Darnley Jr., James D.
; REGISTRATION NUMBER: 33,673
; REFERENCE/DOCKET NUMBER: 4755.P CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 616-385-5210
; TELEFAX: 616-385-6897
; TELEX: 224401
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1617 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; PCT-US94-02552-18

Query Match      8.18; Score 232.2; DB 5; Length 1617;
Best Local Similarity 54.8; Pred. No. 5.2e-40;
Matches 612; Conservative 0; Mismatches 463; Indels 42; Gaps 6;

QY      472  ATTATGATAAATTGCCCCAGGACATCTGTTATCATGACATTTTATAATGAAGCCCTGGTCAA 531
Db      269  ATCCAGATAACCTTCCTACACCAGTGTGGTGATTGTTTCCACAATGAGGCTTGGAGCA 328
QY      532  CTCCTCTCCGAGACAGTTTACAGTGTCTCTGAGACATCCCGGATATCTCTGCTAGAGAAG 591
Db      329  CACTTCTCGAAACTGTCCATAGCGTCATTAATCGCTCACCAAGGCACATGCTAGAGAAA 388
QY      592  TGATCTCTGTAGATGACTACAGTGATAGAGACACCTGAGAGCGCTTCGCCAATGACG 651
Db      389  TTGTTCTTAGTAGATGATGCCAGTGAAGAGACATTTTAAAAAGACCTCTTAGAGAGTTACG 448
QY      652  TTTTCGGGACT---GCCCAAGGTGGCCTGATCCCGCGCAACAAGAGAGAGGGCCTGGTGC 708
Db      449  TGAATAAATAAAGTAGTACCGCTTCACGTCATTGCAATGGAGCAGCGTTCTGGATTGATCA 508
QY      709  GAGCCCGGCTGCTGGGGCGGTCTGCGCGAGGGCGGATGTTCTGACCTTCTCTGACTGTC 768
Db      509  GAGCTAGGTTAAAGGTGCTGCTGTGCTAAAGCCCAAGTCATCACTTTTACGCGCG 568

```



```
REFERENCE/DOCKET NUMBER: 4755.P CP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 616-833-2210
TELEFAX: 616-833-8897
TELEX: 224401
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1680 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cdna
US-08-967-506-11

Query Match      8.1%; Score 232.2; DB 3; Length 1680;
Best Local Similarity 54.8%; Pred. No. 5.3e-40;
Matches 612; Conservative 0; Mismatches 463; Indels 42; Gaps 6;

Qy 472 ATTATGATAATTTGCCAGGACATCTGTTATCATAGCATTTTATAATGAAGCCCTGGTCAA 531
Db 332 ATCCAGATAACCTTCTACACCACTGTTGGTGAATTTTCCACAAATGAGGCTTGGAGCA 391

Qy 532 CTCTCCTTCGGACAGTTTACAGTGTCTTGGAGACATCCCGGATATCTGCTAGAGAAG 591
Db 392 CACTTCTCGCAACTGTCCATAGCGTCAATTAATCGCTCACCAGGACATGCTAGAGAAA 451

Qy 592 TGATCCTGTAGAGTACTACAGTGTATAGAGACACCTGAAGGAGCGCTTGGCCAAATGAGC 651
Db 452 TTGTTCTAGTAGATGATGCCAGTGAAGAGACTTTTAAAGAACCTCTAGAGAGTTACG 511

Qy 652 TTTCCGGACT---GCCCAAGGTGGCCCTGATCCCGGCCAAACAGAGAGAGCGGCTGGTGC 708
Db 512 TGAATAAATAAAGTACCCGTTACGCTATCGAATGGAGCAGCGTTCTGGATTGATCA 571

Qy 709 GAGCCCGCTCTCTGGGGCGTCTCGGCGAGGGCGATGTTCTGACCTTCTGCACTGTC 769
Db 572 GAGCTAGTTAAAGGTGCTGCTGTCTAAAGGCCAAGTGATCACCTTTTACACGCGC 631

Qy 769 ACTGTGATGCCAGNAGGTGGTGGAGCGCTGCTGCAGAGATCCATGAAGAGGT 828
Db 632 ACTGTGATGCCAGTGGGGTGGAGCGCTCTCTTAGCCAGGATCAACATCAGCAGGA 691

Qy 829 CGGAGTGGTGTGCCCGGTGATGATGTGATCGACTGGAAACACCTTGGATACCTGGGGA 888
Db 692 AGACAGTGTCTGTCCCATCATAGATGTGATGATGATGATGATGATGATGATGATGATG 751

Qy 889 ACTCGGGAGGCCAGATCGGGGTTCGACTGGAGGCTGGTGTTCACGTGGGCACACAG 948
Db 752 GTTCTGACATGACCTA---TGGCGGGTTCAACTGGAAGCTCAACTTTTCGCTGTATCTG 808

Qy 949 TTCTCTGAGAGGAGGAGG---ATACGGATGCAATCCCGCTGATGTCATCAGGTCTCAA 1005
Db 809 TTCCCCAAGAGAATGGACAGAGGAAGGTGATCGGACTCTTCTCTGTGAGAACCTTA 868

Qy 1006 CAATGGCTGTGGGCTGTCTCTGTGATGAAGAATAATTTTGAATATCTGGGGTCTTATG 1065
Db 869 CAATGGCAGGAGGCTTTTTCATAGACAGATTAATTTTTCAGAAATTTGGAACATATG 928

Qy 1066 ATACAGGAATGGAATTTGGGAGAGAGAAAACCTCGAATTTCTTTAGGATCTGGCAGT 1125
Db 929 ATGCTGGAATGGATTTTGGGAGAGAGAAAACCTAGAAAATTTCTTTAGGATTTGGCAGT 988

Qy 1126 GTGGTGGGTTCTGGNAACACACCATGTTCCCATGTTGGCCATGTTTTCCTCCCAAGCAAG 1185
Db 989 GTGGAGGAACCTTTGGAGATGTTTACTTGCTCACATGTTGGACATGTTTTCGGAAGCTA 1048

Qy 1186 CTCCTACTCTCCGCAACAGGCTCTGGGCCAACAGTGT-----TCGTGCAG 1230
Db 1049 CACCTACACGTTTCCAGGAGCAGCGGCGAGATTAATTAATAAATAACAGACAGCTTG 1108

Qy 1231 CTGAAGTATGATGAATGAAATTTAAAGAGCTCTACTACCATCGCAACCCCGTGCCTGCT 1290
Db 1109 CAGAAGTATGATGATGAATTTCAAGAAATTTCTTCTATATAATTTCTCCAGGTGTACAA 1168
```

## RESULT 12

PCT-US94-02552-11

; Sequence 11, Application PC/TUS9402552

; GENERAL INFORMATION:

; APPLICANT: Elhammer, Ake P.

; APPLICANT: Homa, Fred L.

; TITLE OF INVENTION: A Cloned DNA Encoding a UDP-GalNac:

; NUMBER OF SEQUENCES: 19

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: The Upjohn Company, Corp. Intellectual

; ADDRESSEE: Property Law

; STREET: 301 Henrietta Street

; CITY: Kalamazoo

; STATE: Michigan

; COUNTRY: USA

; ZIP: 49001

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US94/02552

; FILING DATE:

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Darnley Jr., James D.

; REGISTRATION NUMBER: 33,673

; REFERENCE/DOCKET NUMBER: 4755.P CP

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 616-385-5210

; TELEFAX: 616-385-6897

; TELEX: 224401

; INFORMATION FOR SEQ ID NO: 11:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1680 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: cdna

PCT-US94-02552-11

Query Match 8.1%; Score 232.2; DB 5; Length 1680;

Best Local Similarity 54.8%; Pred. No. 5.3e-40;

Matches 612; Conservative 0; Mismatches 463; Indels 42; Gaps 6;

Qy 472 ATTATGATAATTTGCCAGGACATCTGTTATCATAGCATTTTATAATGAAGCCCTGGTCAA 531

Db 332 ATCCAGATAACCTTCTTACACCAAGTGTGGTATTGTTTCCCAATGAGGCTTGGAGCA 391



Qy	532	CTCTCTTCGACAGTTTTACAGTGTCTTTGAGACATCCCCGGATATCTCTGTAGAAAG	591
Db	392	CACTTCTCGCAACTGTCTCATAGCGTCAATTAATCGCTCACCAAGGCACATGCTAGAGAA	451
Qy	592	TGATCCTTGTAGATGACTACAGTGCATAGAGAGCACTTGAAGGAGCGTTGGCCAATGAGC	651
Db	452	TTGTCTCTAGTAGATGATGCCAGTGAAGAGACATTTTTTAAAGAAGACCTCTAGAGAGTTACG	511
Qy	652	TTTTCGGACT---GCCCAAGGTGGCGCTGATCCGGCCCAACAAGAGAGAGGGCCTGTGTGC	708
Db	512	TGAAAAAATTAAGAGTACCCGTTACGTCATTGCAATGGAGCAGCGTTCTGGATTGATCA	571
Qy	709	GAGCCCGGCTGCTGGGGGCTGTGCGGCGAGGGGCGATGTTCTGCCTTCTCTGGACTGTC	768
Db	572	GAGCTAGGTTAAAAAGGTGCTGCTGTGCTAAAGGCCAAGTGATCACCTTTTACACGCGC	631
Qy	769	ACTGTGAGTGCCACGAAGGCTGGCGCTGAGCGCTGCTGCAGAGAGATCCATGAAGAGAGT	828
Db	632	ACTGTGAGTGACAGTGGGGTGGCTGGAGCGCTCTCTTAGCCAGGATCAAAACATGACAGGA	691
Qy	829	CGGCAGTGTGTGCCCGGTGATTGATGTGATCGACTCGAAACACCTTCGAATACCTGGGGA	888
Db	692	AGACAGTGTGTGCTCCCATCATAGATGTGATCAGTGATGACACTTTGAGTACAATGGCAG	751
Qy	889	ACTCCGGGAGCCCCAGATCGCGGGTTTCGACTGGAGGCTGGTGTTCACGTGGCACACAG	948
Db	752	GTTCTGACATGACCTA--TGGCGGGTTCAACTGGAAGCTCAACTTTTCGCTGGTATCCTG	808
Qy	949	TTCTTGAGAGGGAGAGG---ATACGGATGCAATCCCCGGTGCATGTGCATCAGGTCTCAA	1005
Db	809	TTCCCCAAGAGAGAAATGCACAGAAAGGAAAGGTGATCGGACTCTTCTCTGTGAGAACACCTA	868
Qy	1006	CAATGGCTGTGGGCTGCTTTGCTGTGTGAGTAAGAAATTTTGAATATCTGGGGTCTTATG	1065
Db	869	CAATGGCAGGAGGCGCTTTTTCATAGACAGAGATTACTTTTCAGGNAATTTGGAACATATG	928
Qy	1066	ATACAGGAATCGAAGTTTGGGGAGGAGAAACCTCGAAATTTTCCTTTAGGATCTGGCAGT	1125
Db	929	ATGCTGGAATGGATATTTGGGGAGGAGAAACCTAGAAATTTCTTTAGGATTTGGCAGT	988
Qy	1126	GTGGTGGGTTCTGGAACAACACCCATGTTCCCATGTGGCCATGTTTTCCCAAGCAAG	1185
Db	989	GTGGAGAACTTTGGAGATTGTTACTTGTCTCACATGTTGGACATGTGTTTCGGAAGCTA	1048
Qy	1186	CTCCCTACTCCGCAACAAGGCTCTGGCCACAGTGT-----TGTGTCAG	1230
Db	1049	CACCTTACAGTTTCCAGSAGGCACGGGCGAGATTATCAATAAAAAATACAGACGACTTG	1108
Qy	1231	CTGAAGTATGGATGGATGAATTTAAAGAGCTCTACTACCATCGCAACCCCGTCGCCGCT	1290
Db	1109	CAGAAATATGGATGGATGAATTCAGAAATTTCTTCTATATAATTTCTCCAGGTGTTACAA	1168
Qy	1291	TGGAACCTTTTGGGATGTGACAGAGAGGAAGCAGCTCCGGGCAAGCTCCAGTGTAAAG	1350
Db	1169	AGGTAGATTATGGAGATATATATCAAGACTTTGGTCTTAAGSCACAAACTCCAATGCAGAC	1228
Qy	1351	ACTTCAAGTGGTCTTCGAGACTGTGTATCCAGAACTGCATGTGCCTGAGAGACAGGCTG	1410
Db	1229	CATTCTCTTGGTACCTAGAGAAATTTATCCTGATTTCTCAGATTTCTTCGTCACATATTTCT	1288
Qy	1411	GCTTTCTCGGATGCTCCAGAAACAAGGACTAACAGACTACTGCTTTTGACTATAACCCCTC	1470
Db	1289	CTTTGGAGAGATAC-----GAAATGTGAAACAATACTAGTGTCTAGATAACATGGCTA	1342
Qy	1471	CCGATGAAAAACAGATTGTGGGACACACAGGTCAATTTCTGTACCTCTGTATGGGATGGGCC	1530
Db	1343	GAAGAAGAGATGAAGAATTTGGA-----ATTTTAACTGTCTAGTATGGGAG	1390
Qy	1531	AGAATCAGTTTTTCAGTACACGTCCTCCAGAAAGAAAT	1567
Db	1391	GTAAATCAGGTTTTCTCTTACACTGCCAACCAAGAAAT	1427

```

RESULT 13
US-08-967-508-10
; Sequence 10, Application US/08967508
; Patent No. 5910570
; GENERAL INFORMATION:
; APPLICANT: The Upjohn Company
; APPLICANT: FOR U.S. PURPOSES ONLY: Elhammer, Ake P. and Homa, Fred L.
; TITLE OF INVENTION: A cloned DNA Encoding a UDP-GalNAC:
; TITLE OF INVENTION: Polypeptide, N-Acetyl-galactosaminyltransferase
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pharmacia & Upjohn Company, Intellectual
; ADDRESSEE: Property Legal Services
; STREET: 301 Henrietta Street
; CITY: Kalamazoo
; STATE: Michigan
; COUNTRY: USA
; ZIP: 49001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/967,508
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/602,830
; FILING DATE: 13 No. 5910570ember 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Darnley Jr., James D.
; REGISTRATION NUMBER: 33,673
; REFERENCE/DOCKET NUMBER: 4755.P CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 616-833-2210
; TELEFAX: 616-833-8897
; TELEX: 224401
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2294 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-967-508-10

Query Match      8.1%; Score 232.2; DB 2; Length 2294;
Best Local Similarity 54.8%; Pred. No. 6.1e-40;
Matches 612; Conservative 0; Mismatches 463; Indels 42; Gaps 6;

Qy 472 ATTATGATAATTGCCAGACATCTGTTATCATAGCATTTTATATGAAGCCTGGTCAA 531
Db 384 ATCAGATAACCTTCCTACCAACGAGTGGTGATGTTTTCACAAATGAGGCTTGGAGCA 443

Qy 532 CTCTCTTCGGACAGTTTACAGTGTCTCTTGAGACATCCCGCGATATCCTGCTAGAAGAAG 591
Db 444 CACTTCTCGAACTGTCCATAGCGTCATTATCGCTCACCACGGCACATGCTAGAAGAA 503

Qy 592 TGATCTCTTGATAGTACTACAGTGATAGAGACACCTGAAGAGGCGCTTGGCCAAATGAGC 651
Db 504 TTGTTCTAGTAGATGATGCCAGTGAAGAGACTTTTAAAGACCTCTAGAGAGTTACG 563

Qy 652 TTTTCGGGACT--GCCCAAGTGGCTGATCGCGCCCAACAGAGAGAGGGCTGTGTGC 708
Db 564 TGAATAAATTAAGTAGTACCGGTTTCAGTCATTGGAATGGAGACGCTTCTGGATTGATCA 623

Qy 709 GAGCCCGCTGCTGGGGCGCTCTGCGCGAGGGCGGATGTTCTGACCTTCTTGGACTGTC 768
Db 624 GAGCTAGTTAAAGGTGCTGCTGTCTTAAGGCCAAGTGATCACCTTTTAGACGGC 683

Qy 769 ACTGTGAGTGCCACGAAGGGTGGCTGGAGCGCTGCTGTCAGAGGATCCATGAAGAGGAGT 828

```



```
QY 1066 ATACAGGATGGAAGTTTGGGAGGAGAAACCTCGAATTTCTTTAGGATCTGGCAGT 1125
D 981 ATGCTGGAATGGATATTTGGGAGGAGAAACCTAGAAATTTCTTTAGGATTTGGCAGT 1040
QY 1126 GTGCTGGGGTCTCGAAACACACACCATGTTCCCATGTTGGCATGTTTCCCAAGCAAG 1185
D 1041 GTGGAGGAACCTTGGAGATTTGTTACTTGTCTCACATGTTGACATGTTTTCGGAAAGCTA 1100
QY 1186 CTCCTACTCCCGCAACAAAGCTCTGGCCACACAGTGT-----TCGTGCAG 1230
D 1101 CACCTACACGTTTCCAGGAGGCGACGGGCGAGATTATCAATAAAAAATAACAGACGACTTG 1160
QY 1231 CTGAAGTATGATGATGATGATTTAAAGAGCTCTACTACCATCGCAACCCCGTCCCGCT 1290
D 1161 CAGAAGTATGATGATGATGATTTCAAGAAATTTCTTATATAATTTCTCCAGGTTTACAA 1220
QY 1291 TGAACCTTTTGGGGATGTGACAGAGAGGAAGCAGCTCCGGGCAAGCTCCAGGTGTAAG 1350
D 1221 AGGTAGATATGGAGATATATCATCAAGACTTGGTCTAAGGCACAACTCCAATGCAGAC 1280
QY 1351 ACTTCAAGTGGTCTTGGAGACTGTGTATCCAGAACTGTCATGTCCTGAGGACAGGCTG 1410
D 1281 CATCTCTGCTGCTAGAGAAATTTTATCTGATTTCTCAGATTCTCGTCACTATTCTTCT 1340
QY 1411 GCTTCTTGGGATGCTCCAGAACAAAGACTTAACAGACTACTGTTTGCATTAACCTC 1470
D 1341 CTTTGGAGAGATAC-----GAAATGTGGAACAAATCATGTTCTAGATAAATGCTA 1394
QY 1471 CCGATGAAACCAAGATTGTGGGACACCCAGGTCACTTCTGTACCTCTGTGATGGGATGGCC 1530
D 1395 GAAAGAGATGAAAGTTGGA-----ATTTTAACTGTCTATGGTATGGAG 1442
QY 1531 AGAATCAGTTTTCGAGTACACGTCCAGAAAGAAAT 1567
D 1443 GTAATCAGTTTCTCTTACACTGCCAACAAAGAAAT 1479
```

## RESULT 15

```
PCT-US94-02552-10
; Sequence 10, Application PC/TUS9402552
; GENERAL INFORMATION:
; APPLICANT: Elhammer, Ake P.
; APPLICANT: Homa, Fred L.
; TITLE OF INVENTION: A Cloned DNA Encoding a UDP-GalNac:
; TITLE OF INVENTION: Polypeptide, N-Acetylgalactosaminyltransferase
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Upjohn Company, Corp. Intellectual
; ADDRESSEE: Property Law
; STREET: 301 Henrietta Street
; CITY: Kalamazoo
; STATE: Michigan
; COUNTRY: USA
; ZIP: 49001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/02552
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Darnley Jr., James D.
; REGISTRATION NUMBER: 33,673
; REFERENCE/DOCKET NUMBER: 4755.P CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 616-385-5210
; TELEFAX: 616-385-6897
; TELEX: 224401
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
```

```
; LENGTH: 2294 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; PCT-US94-02552-10
```

```
Query Match 8.1%; Score 232.2; DB 5; Length 2294;
Best Local Similarity 54.8%; Pred. No. 6.1e-40;
Matches 612; Conservative 0; Mismatches 463; Indels 42; Gaps 6;
```

```
QY 472 ATTATGATATTTGCCAGGACATCTGTATCATATAGCAATTTTATAATGAACCTTGGTCAA 531
D 384 ATCCAGATAACCTTCTTACAAACAGTGTGGTGATTTTCCCAATGAGGCTTGGAGCA 443
QY 532 CTCTCTCTCGGACAGTTTACAGTGTCTTCAGACATCCCGGATATCTCTGCTAGAAAG 591
D 444 CATCTCTCGAACTGTCCATAGCGTCATTAATCGCTCACCAGGCATCATCTAGAGAAA 503
QY 592 TGATCTCTTGTAGTACTACAGTGTAGAGAGACCTGAAAGAGCGCTTGGCCAAATGAGC 651
D 504 TTGTTCTAGTAGATGATGCGAGTGAAGAGACTTTTAAAGAGACCTCTAGAGATTACG 563
QY 652 TTTGGGGACT--GCCCCAAGTGGCTGTATCCCGGCAACAAGAGAGAGGCGCTGTGTC 708
D 564 TGAATAAATTTAAAGATACCCGTTTACGTCTCATTCGAATGGAGCAGCGTTCTGGATTGATCA 623
QY 709 GAGCCCGGCTGTCTGGGGCGTCTCGCGGAGGGCGGATGTTCTGACCTTCTCTGACTGTC 768
D 624 GAGCTAGTTTAAAGTGTCTGTCTTAAGGCCAAGTATGATCACTTTTATAGACCGC 683
QY 769 ACTGTGAGTGCACGAAGGTGTCTGGAGCGCTGTCTGACAGAGATCATGAAGAGAGT 828
D 684 ACTGTGAGTGCACAGTGGGGTGTCTGAGCGCTCTCTTAGCCAGGATCAAAACATGACAGGA 743
QY 829 CGGAGTGTGTCTCCCGGTGATGTGATGTGATGTGATGTGATGTGATGTGATGTGATGTG 888
D 744 AGACAGTGTGTCTCTCCCATCATAGATGTGATGTGATGTGATGTGATGTGATGTGATGTG 803
QY 889 ACTCCGGGAGCCCGAGATCGCGGTCTCGAGTGGAGGCTGGTGTTCACCTGGGCACACAG 948
D 804 GTTCTGACATGACCTA---TGGCGGGTTCACTTGGAGCTCAACTTTCGCTGATCTCTG 860
QY 949 TTCTTGGAGGGGAGAGG---ATACGGATGCAATCCCCCGTCGATGTGATCATCAGGTCTCAA 1005
D 861 TTCCCAAGAGAGAAATGGACAGAAAGGAGTGTGATCGGACTCTTCTGTGAGAAACCTA 920
QY 1006 CAATGCTGTGGGCTGTGTCTGTGATGATGATGATGATGATGATGATGATGATGATGATG 1065
D 921 CAATGCGAGGAGCGCTTTTTCATATAGACAGAGATTACTTTCAGGAAATTTGGAACATATG 980
QY 1066 ATACAGGAATGGAAGTTTGGGAGGAGAAACCTCGAATTTTCTTTAGGATCTGGCAGT 1125
D 981 ATGCTGGAATGGATATTTGGGAGGAGAAACCTAGAAATTTCTTTAGGATTTGGCAGT 1040
QY 1126 GTGGTGGGGTCTTGAAACACACACCATGTTCCCATGTTGGCCATGTTTCCCAAGCAAG 1185
D 1041 GTGGAGGAACCTTGGAGATTTGTTACTTGTCTACATGTGACATGTTTTCGGAAGCTA 1100
QY 1186 CTCCTACTCCCGCAACAAAGCTCTCGGCAACAGTGT-----TCGTGCAG 1230
D 1101 CACCTTACACGTTTCCAGGAGGCGACGGGCGAGATTATCAATAAAAAATAACAGACGACTTG 1160
QY 1231 CTGAAGTATGATGATGATTTAAAGAGCTCTACTACCATCGCAACCCCGTCCCGCT 1290
D 1161 CAGAAGTATGATGATGATTTCAAGAAATTTCTTATATAATTTCTCCAGGTTTACAA 1220
QY 1291 TGAACCTTTTGGGGATGTGACAGAGAGGAAGCAGCTCCGGGCAAGCTCCAGGTGTAAG 1350
D 1221 AGGTAGATATGGAGATATATCATCAAGACTTGGTCTAAGGCACAACTCCAATGCAGAC 1280
QY 1351 ACTTCAAGTGGTCTTGGAGACTGTGTATCCAGAACTGTCATGTCCTGAGGACAGGCTG 1410
```

Db 1281 CATTCTCTTGTAAGAGATATTTATCTGATTCAGATTCCTCGTCACTATTCT 1340  
Qy 1411 GCTTCTTCGGGATGCTCCAGNACAAAGACTACAGACTACTGCTTTGACTATAACCTC 1470  
Db 1341 CTTTGGGAGAGATAC-----GAAATGTGGAACAAATCAGTGTCTAGATAACATGGCTA 1394  
Qy 1471 CCGATGAAACACGATTGTGGGACACCAGGTCAATCTGTACCTCTGTATGGGATGGCC 1530  
Db 1395 GAAAGAGATGAAAGTTGGA-----ATTTTAACTGTATGGGATGGGAG 1442  
Qy 1531 AGAATCAGTTTTTCGAGTACACGTCGCCAGAAAGAAAT 1567  
Db 1443 GTAATCAGGTTTTTCTCTTACACTGCCAACAAAGAAAT 1479

Search completed: December 7, 2003, 21:21:29  
Job time : 180.847 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 7, 2003, 18:38:04 ; Search time 867.526 Seconds  
(without alignments)  
10918.700 Million cell updates/sec

Title: US-10-074-527-1  
Perfect score: 2850  
Sequence: 1 cggcgctggtaccactataa.....aaaaaaaaaaaaaaaaaaaaa 2850

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2201672 seqs, 1661799599 residues

Total number of hits satisfying chosen parameters: 4403344

Minimum DB seq length: 0  
Maximum Match 100%

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA.\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*
- 14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*
- 16: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*
- 17: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2848	99.9	2850	13	US-10-074-527-1
2	2802.4	98.3	2904	14	US-10-198-846-9749
3	1746	61.3	1746	13	US-10-074-527-3
4	1744.4	61.2	1746	12	US-10-292-896-57
5	1208.4	42.4	1259	9	US-09-925-299-219
6	1208.4	42.4	1259	11	US-09-925-299-219
7	1012.2	35.5	1517	14	US-10-106-698-598
8	1011.8	35.5	1517	11	US-09-895-298-44
9	584.6	20.5	818	14	US-10-198-846-7621
10	526.4	18.5	544	9	US-09-777-564-414
11	526.4	18.5	544	14	US-10-015-219-414
12	524	18.4	525	12	US-10-292-896-121
13	496.4	17.4	632	9	US-09-777-564-172
14	496.4	17.4	632	9	US-09-777-564-1423
15	496.4	17.4	632	14	US-10-015-219-172
16	496.4	17.4	632	14	US-10-015-219-1423

17	450.8	15.8	496	10	US-09-867-701-4790	Sequence 4790, Ap
18	443.4	15.6	473	10	US-09-867-701-4789	Sequence 4789, Ap
19	425.8	14.9	473	10	US-09-867-701-3683	Sequence 3683, Ap
20	409	14.4	421	10	US-09-967-768A-21	Sequence 21, Appl
21	409	14.4	421	12	US-09-873-367C-550	Sequence 550, App
22	406.4	14.3	408	10	US-09-867-701-10399	Sequence 10399, A
23	285.6	10.0	1920	9	US-09-789-417-1	Sequence 1, Appli
24	285.6	10.0	1920	9	US-09-792-451-1	Sequence 1, Appli
25	285.6	10.0	1920	11	US-10-292-896-59	Sequence 59, Appl
26	285.6	10.0	2575	11	US-09-946-374-346	Sequence 346, App
27	285.6	10.0	2575	12	US-10-015-387A-346	Sequence 346, App
28	285.6	10.0	2575	12	US-10-006-130A-346	Sequence 346, App
29	285.6	10.0	2575	12	US-10-199-672-363	Sequence 363, App
30	285.6	10.0	2575	12	US-10-006-172A-346	Sequence 346, App
31	285.6	10.0	2575	12	US-10-187-749-363	Sequence 363, App
32	285.6	10.0	2575	12	US-10-194-457-363	Sequence 363, App
33	285.6	10.0	2575	12	US-10-184-642-363	Sequence 363, App
34	285.6	10.0	2575	12	US-10-196-747-363	Sequence 363, App
35	285.6	10.0	2575	12	US-10-015-392A-346	Sequence 346, App
36	285.6	10.0	2575	12	US-10-017-253A-346	Sequence 346, App
37	285.6	10.0	2575	12	US-10-173-689-363	Sequence 363, App
38	285.6	10.0	2575	12	US-10-173-690-363	Sequence 363, App
39	285.6	10.0	2575	12	US-10-173-691-363	Sequence 363, App
40	285.6	10.0	2575	12	US-10-173-692-363	Sequence 363, App
41	285.6	10.0	2575	12	US-10-173-694-363	Sequence 363, App
42	285.6	10.0	2575	12	US-10-173-698-363	Sequence 363, App
43	285.6	10.0	2575	12	US-10-173-699-363	Sequence 363, App
44	285.6	10.0	2575	12	US-10-173-707-363	Sequence 363, App
45	285.6	10.0	2575	12	US-10-174-569-363	Sequence 363, App

## ALIGNMENTS

## RESULT 1

US-10-074-527-1  
; Sequence 1, Application US/10074527  
; Publication No. US20020142426A1  
; GENERAL INFORMATION:  
; APPLICANT: Olandt, Peter J.  
; APPLICANT: Meyers, Rachel E.  
; APPLICANT: Galvin, Katherine A.  
; APPLICANT: Millennium Pharmaceuticals Inc.  
; TITLE OF INVENTION: 33945, A Human Glycosyltransferase and  
; FILE REFERENCE: MPI2001-018P1RCP1(M)  
; CURRENT APPLICATION NUMBER: US/10/074,527  
; CURRENT FILING DATE: 2002-02-12  
; PRIOR APPLICATION NUMBER: 60/269202  
; PRIOR FILING DATE: 2001-02-15  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: Fast-Seq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 2850  
; TYPE: DNA  
; ORGANISM: homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (81)....(1826)  
US-10-074-527-1

Query Match 99.9%; Score 2848; DB 13; Length 2850;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2850; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	CCGGCTCGGTACCACTATAACGCCGCCAGTGTGCTGGAATTCGCCCTTCGCAGATCGC	60
Db	1	CCGGCTCGGTACCACTATAACGCCGCCAGTGTGCTGGAATTCGCCCTTCGCAGATCGC	60
Qy	61	TGCTCGACTTGGCGGGCGGCATGTGGGGCGCACCGCGCGCGCTGCCCGGGGAAC	120
Db	61	TGCTCGACTTGGCGGGCGGCATGTGGGGCGCACCGCGCGCGCTGCCCGGGGAAC	120

121 TGCGGCGCGCGCGCGCGCGCTGTGTGTCTCTGCGCTACTGCGCTTGGCGGCTGG 180  
121 TGCGGCGCGCGCGCGCGCGCTGTGTGTCTCTGCGCTACTGCGCTTGGCGGCTGG 180  
181 GCTCGGTGCTCGCGGCGCGAGCGTGGGCGCGGCGCGGCGTGCAGAGCGCGGCGCG 240  
181 GCTCGGTGCTCGCGGCGCGAGCGTGGGCGCGGCGCGGCGTGCAGAGCGCGGCGCG 240  
241 GCACCCGCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300  
241 GCACCCGCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300  
301 CGCTGGGCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360  
301 CGCTGGGCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360  
361 AGAGGAGAGCGTGCCTGCAACAGATTAACATCTACCTACGCGACCGCATCTCACTGC 420  
361 AGAGGAGAGCGTGCCTGCAACAGATTAACATCTACCTACGCGACCGCATCTCACTGC 420  
421 ACCGCGCTGCGCGAGCGTGGAAACCGCTGTGCAAGAGAGAAATATGATATGATA 480  
421 ACCGCGCTGCGCGAGCGTGGAAACCGCTGTGCAAGAGAGAAATATGATATGATA 480  
481 ATTTGCCAGAGACATCTGTTTATCATAGCATTTTATATGAGCTGCTCACTCTCTTC 540  
481 ATTTGCCAGAGACATCTGTTTATCATAGCATTTTATATGAGCTGCTCACTCTCTTC 540  
541 GGACAGTTTACAGTGTCTTTGAGACATCCCGGATATCTCTGCTAGAGAAAGTATCTTG 600  
541 GGACAGTTTACAGTGTCTTTGAGACATCCCGGATATCTCTGCTAGAGAAAGTATCTTG 600  
601 TAGATGACTACAGTATAGAGAGACCTGAGAGCGCTTGGCCATGAGCTTTCGGGAC 660  
601 TAGATGACTACAGTATAGAGAGACCTGAGAGCGCTTGGCCATGAGCTTTCGGGAC 660  
661 TGCCCAAGGTGCGCTCATCCGCGCAACAGAGAGAGCGCTGCTGCGAGCGCGGCTGC 720  
661 TGCCCAAGGTGCGCTCATCCGCGCAACAGAGAGAGCGCTGCTGCGAGCGCGGCTGC 720  
721 TGGGGCGCTGCGCGCGAGGGCGATGTTCTGACCTTCTTGACTGTCACTGTGAGTGC 780  
721 TGGGGCGCTGCGCGCGAGGGCGATGTTCTGACCTTCTTGACTGTCACTGTGAGTGC 780  
781 ACNAGGGTGGCTGGAGCGCTGCTGAGAGGATCCATGAGAGAGTGGCAGTGGTGT 840  
781 ACNAGGGTGGCTGGAGCGCTGCTGAGAGGATCCATGAGAGAGTGGCAGTGGTGT 840  
841 GCGCGGTGATTGATGATCGACTGGAAACCTTTCGAATACCTGCGGAACTCCCGGGAGC 900  
841 GCGCGGTGATTGATGATCGACTGGAAACCTTTCGAATACCTGCGGAACTCCCGGGAGC 900  
901 CCCAGATCGGCGGTTTCGACTGGAGGCTGTTTCACTGCGCACACAGTTCCTGAGAGG 960  
901 CCCAGATCGGCGGTTTCGACTGGAGGCTGTTTCACTGCGCACACAGTTCCTGAGAGG 960  
961 AGAGGATACGATGCAATCCCGCTGATGCTCAGTCTCAGTCTCAGTCTCAGTCTGCGC 1020  
961 AGAGGATACGATGCAATCCCGCTGATGCTCAGTCTCAGTCTCAGTCTCAGTCTGCGC 1020  
1021 TGTGTTGCTGAGTAAGAAATATTTGAAATCTGCGGCTTATGATACAGGAATGGAAG 1080  
1021 TGTGTTGCTGAGTAAGAAATATTTGAAATCTGCGGCTTATGATACAGGAATGGAAG 1080  
1081 TTTGGGAGGAGAAACCTCGAAATTTCTTTAGGATCTGCGAGTGTGGGTTCTGG 1140  
1081 TTTGGGAGGAGAAACCTCGAAATTTCTTTAGGATCTGCGAGTGTGGGTTCTGG 1140  
1141 AAACACACCCATGTTCCATGTTGGCATGTTTTTCCCGAAGCAAGCTCCCTACTCCCGCA 1200  
1141 AAACACACCCATGTTCCATGTTGGCATGTTTTTCCCGAAGCAAGCTCCCTACTCCCGCA 1200  
1201 ACNAGGCTCTGGCCAAACAGTGTCTGTCAGCTGAAGTATGATGATGATTTAAAGAGC 1260

1201 ACAAGGCTCTGGCCAAACAGTGTTCGTGAGCTGAAGTATGATGATGATTTAAAGAGC 1260  
1261 TCTACTACCATCGCAACCCCGTGGCGCTTGGAACTTTTGGGGATGTCAGAGAGGA 1320  
1261 TCTACTACCATCGCAACCCCGTGGCGCTTGGAACTTTTGGGGATGTCAGAGAGGA 1320  
1321 AGCAGCTCCGGGACAAAGCTCCAGTGTAAAGACTTCAAGTGGTTCCTTGGAGACTGTGTATC 1380  
1321 AGCAGCTCCGGGACAAAGCTCCAGTGTAAAGACTTCAAGTGGTTCCTTGGAGACTGTGTATC 1380  
1381 CAGAACTGATGTGCTGAGGACAGGCTGGCTTCTTGGGATGCTCCAGAAACAAAGGAC 1440  
1381 CAGAACTGATGTGCTGAGGACAGGCTGGCTTCTTGGGATGCTCCAGAAACAAAGGAC 1440  
1441 TAACAGACTACTGCTTGTGACTATAACCTCCCGATGAACACCAAGATTTGGGACACAGG 1500  
1441 TAACAGACTACTGCTTGTGACTATAACCTCCCGATGAACACCAAGATTTGGGACACAGG 1500  
1501 TCATTCTGTACCTCTGCTGATGGGATGGGCGCAGAAATCAGTTTTTTCGAGTACACGTCGCCAGA 1560  
1501 TCATTCTGTACCTCTGCTGATGGGATGGGCGCAGAAATCAGTTTTTTCGAGTACACGTCGCCAGA 1560  
1561 AAGAAATACGCTATAACACCCAGCTGAGGCTGCATTTGCTGTGGAAAGCAGGAATGG 1620  
1561 AAGAAATACGCTATAACACCCAGCTGAGGCTGCATTTGCTGTGGAAAGCAGGAATGG 1620  
1621 ATACCCTTATCATGCTCTCTGCAAGAAATCTGCCCGCAGAGAAATCAGAACTTCACTTGC 1680  
1621 ATACCCTTATCATGCTCTCTGCAAGAAATCTGCCCGCAGAGAAATCAGAACTTCACTTGC 1680  
1681 AGGAGGATGATCTTTTATTTTACGAAACAGCTGCAAGAAATGCTGCCAGGCTCGAGGAAGG 1740  
1681 AGGAGGATGATCTTTTATTTTACGAAACAGCTGCAAGAAATGCTGCCAGGCTCGAGGAAGG 1740  
1741 AGTCGAGTGCACAGTTCGTTTCCACTCTTACGAGACTGCACCAACTCGGATCATCAGAAAT 1800  
1741 AGTCGAGTGCACAGTTCGTTTCCACTCTTACGAGACTGCACCAACTCGGATCATCAGAAAT 1800  
1801 GGTTCCTTCAAGAGCGATGTTATGAGCTCTGTATCAAGGAGCCCATCGAAGGAGAC 1860  
1801 GGTTCCTTCAAGAGCGATGTTATGAGCTCTGTATCAAGGAGCCCATCGAAGGAGAC 1860  
1861 TGTGGAGCCAGGACTCTGCGCAACAAAGACTTAGCTAAGCAGTGCACAGAACCCACCAAA 1920  
1861 TGTGGAGCCAGGACTCTGCGCAACAAAGACTTAGCTAAGCAGTGCACAGAACCCACCAAA 1920  
1921 AACTAGGCTGCATTTGCTTTGAAGAGGCAATCAATTTGCGCATTTGTGAAAGTGTGTGGA 1980  
1921 AACTAGGCTGCATTTGCTTTGAAGAGGCAATCAATTTGCGCATTTGTGAAAGTGTGTGGA 1980  
1981 TTTAGTAAATGATGATAGCTTTGTACTTATTTTGAAGACTTTTAAATGTTTCCAAA 2040  
1981 TTTAGTAAATGATGATAGCTTTGTACTTATTTTGAAGACTTTTAAATGTTTCCAAA 2040  
2041 TACCTATTTTCAAGGCTAATCGTAAAGTGTAAACCTTTGGTATTTAGAAAATTTAAAC 2100  
2041 TACCTATTTTCAAGGCTAATCGTAAAGTGTAAACCTTTGGTATTTAGAAAATTTAAAC 2100  
2101 CTTATAATATTTTCTATCAARAWRWAATTTTACAGTCTGCGCTTTTACTCTCATTAGC 2160  
2101 CTTATAATATTTTCTATCAARAWRWAATTTTACAGTCTGCGCTTTTACTCTCATTAGC 2160  
2161 AAAAAGATAAAGATTTTATTTTGGTATTTTACAAGAAATCCAGGATCAGGAATATCTGC 2220  
2161 AAAAAGATAAAGATTTTATTTTGGTATTTTACAAGAAATCCAGGATCAGGAATATCTGC 2220  
2221 ATGCTGGAATCAGGTTCAAGCAACGCTCTTTGCTAATTAAGTAACTACCTCAGCTGCG 2280  
2221 ATGCTGGAATCAGGTTCAAGCAACGCTCTTTGCTAATTAAGTAACTACCTCAGCTGCG 2280  
2281 GGGTAAAGTTCCTCCAGTATAGAGACTGTCTCACTAGGAACATTTGTTATGATTTTCA 2340

Db 2281 GGGTTAAAGTTTCCCAAGTATAGAGACTGTCTACTAGGACATGTATTGATTATTC 2340  
Qy 2341 GGTCAITGAGATCTCTAGATGTAATTTTAAAGAAATGCTTTTGGTGTATGTGTCTAC 2400  
Db 2341 GGTCAITGAGATCTCTAGATGTAATTTTAAAGAAATGCTTTTGGTGTATGTGTCTAC 2400  
Qy 2401 CACAGTTAACTCCATTAATGTTTCATGTGAGCCAAAGAGACTAACCAAGCTGAATCT 2460  
Db 2401 CACAGTTAACTCCATTAATGTTTCATGTGAGCCAAAGAGACTAACCAAGCTGAATCT 2460  
Qy 2461 CAGAGAAATTTGCTTTTACTAGCTCAGTCACTTCTGAGAGCGAATCTTAAACAATGCCG 2520  
Db 2461 CAGAGAAATTTGCTTTTACTAGCTCAGTCACTTCTGAGAGCGAATCTTAAACAATGCCG 2520  
Qy 2521 CACTGTAGTGTGCTGTTCTTACACATGATGACTTTTAAACATGTTTATATCATTTTAAAT 2580  
Db 2521 CACTGTAGTGTGCTGTTCTTACACATGATGACTTTTAAACATGTTTATATCATTTTAAAT 2580  
Qy 2581 TTTTATGATACGGTAGTGTGAGGAGAAATGTAATGTTCTATATGAAATTCCTTTTCAA 2640  
Db 2581 TTTTATGATACGGTAGTGTGAGGAGAAATGTAATGTTCTATATGAAATTCCTTTTCAA 2640  
Qy 2641 GTTGTGTTTCAATTAACAGATTATTAATTTAAATCAGCGTTAGAGTTTGTCTGCTGCAAC 2700  
Db 2641 GTTGTGTTTCAATTAACAGATTATTAATTTAAATCAGCGTTAGAGTTTGTCTGCTGCAAC 2700  
Qy 2701 TGCTGTGAAATTTCTCTGAGTAAATCTGATTTTGTGAATGATCCAGACCAACCTTGAGA 2760  
Db 2701 TGCTGTGAAATTTCTCTGAGTAAATCTGATTTTGTGAATGATCCAGACCAACCTTGAGA 2760  
Qy 2761 TTTTGTCAACCTGATTAAGTCAATATGAATGATTTAAAGATGTGAGAACAAAAA 2820  
Db 2761 TTTTGTCAACCTGATTAAGTCAATATGAATGATTTAAAGATGTGAGAACAAAAA 2820  
Qy 2821 AAAAAAAAAAAAAAAAAAAAAAAAAA 2850  
Db 2821 AAAAAAAAAAAAAAAAAAAAAAAAAA 2850

RESULT 2  
US-10-198-846-9749/c  
; Sequence 9749, Application US/10198846  
; Publication No. US2003009974A1  
; GENERAL INFORMATION:  
; APPLICANT: Lillie, James  
; APPLICANT: Xu, Yongyao  
; APPLICANT: Wang, Youzhen  
; APPLICANT: Steimann, Kathleen  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS  
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
; FILE REFERENCE: MRI-049  
; CURRENT APPLICATION NUMBER: US/10/198,846  
; PRIOR FILING DATE: 2002-07-18  
; PRIOR APPLICATION NUMBER: 60/306,220  
; PRIOR FILING DATE: 2001-07-18  
; NUMBER OF SEQ ID NOS: 14084  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9749  
; LENGTH: 2904  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 1, 2, 3, 4, 5, 2903, 2904  
; OTHER INFORMATION: n = A,T,C or G  
US-10-198-846-9749

Query Match 98.3%; Score 2802.4; DB 14; Length 2904;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 2811; Conservative 1; Indels 1; Gaps 1;  
Qy 12 CCCTATAACGGCGCGCAGTGTCTGGAATTCGCCCTTCGCAGATGCTGTCAGTT 71

Db 2894 CCCTATAACGGCGCGC -GTGTCTGGAATTCGCCCTTCGCAGATGCTGTCAGTT 2836  
Qy 72 GGGGGCGCATGTGGGGCGCACGGCGCGCGCTGCCCGGGGAACCTGGCGCGGC 131  
Db 2835 GGGGGCGCATGTGGGGCGCACGGCGCGCGCTGCCCGGGGAACCTGGCGCGGC 2776  
Qy 132 CGGAGGCGCTGTGTGTCTCTGGGGCTACTGGCGCTTGGCGGGCTGGGCTCGTCTG 191  
Db 2775 CGGAGGCGCTGTGTGTCTCTGGGGCTACTGGCGCTTGGCGGGCTGGGCTCGTCTG 2716  
Qy 192 CGGCGGAGGCTGGGGCGGGGCGGGGCTGCCAGCGCGGACCCCGCGCACCCCGCGC 251  
Db 2715 CGGCGGAGGCTGGGGCGGGGCGGGGCTGCCAGCGCGGACCCCGCGCACCCCGCGC 2656  
Qy 252 CCGGCGGCGCGGCGCGGCTCATGCCGCGCGCGCTGCCAGCGCGGACCCCGCGCACCCCGCGC 311  
Db 2655 CCGGCGGCGCGGCGCGGCTCATGCCGCGCGCGCTGCCAGCGCGGACCCCGCGCACCCCGCGC 2596  
Qy 312 CGGCGGAGGCGGCTGGGGCTGCAGCTGCAGGGCGAGGAGCTGGGGCTGCAGGAGGAGC 371  
Db 2595 CGGCGGAGGCGGCTGGGGCTGCAGCTGCAGGGCGAGGAGCTGGGGCTGCAGGAGGAGC 2536  
Qy 372 GTGCGGCTGCACAGATTAACTACTCAGCGAACCGCATCTCATCTGCACCGCGCGCTG 431  
Db 2535 GTGCGGCTGCACAGATTAACTACTCAGCGAACCGCATCTCATCTGCACCGCGCGCTG 2476  
Qy 432 CCGGAGGCTGGAAACCGCTGTGCAAGAGAGAAATATGATTATGATAATTTGCCCAGG 491  
Db 2475 CCGGAGGCTGGAAACCGCTGTGCAAGAGAGAAATATGATTATGATAATTTGCCCAGG 2416  
Qy 492 ACATCTGTTATCATAGCATTTTATAATGAAGCCTGGTCAACTCTCTCCGAGAGTTTAC 551  
Db 2415 ACATCTGTTATCATAGCATTTTATAATGAAGCCTGGTCAACTCTCTCCGAGAGTTTAC 2356  
Qy 552 AGTGTCTTGAAGCATCCCGGATATCTGCTAGAGAGAGATGATCTCTTGTAGATGACTAC 611  
Db 2355 AGTGTCTTGAAGCATCCCGGATATCTGCTAGAGAGAGATGATCTCTTGTAGATGACTAC 2296  
Qy 612 AGTGTAGAGAGCAGCTGGAAGGAGCGCTTGGCCAAATGAGCTTTCGGGACTGCCCAAGTG 671  
Db 2295 AGTGTAGAGAGCAGCTGGAAGGAGCGCTTGGCCAAATGAGCTTTCGGGACTGCCCAAGTG 2236  
Qy 672 CGCTGATCCGCGCAACAGAGAGAGGCGCTGCTCGAGCGCGGCTGCTGGGGGCGCTCT 731  
Db 2235 CGCTGATCCGCGCAACAGAGAGAGGCGCTGCTCGAGCGCGGCTGCTGGGGGCGCTCT 2176  
Qy 732 CGGCGAGGGCGATGTTCTGACCTTCTGGAATGCTCACTGTAGTGCCACGAAGGGTGG 791  
Db 2175 CGGCGAGGGCGATGTTCTGACCTTCTGGAATGCTCACTGTAGTGCCACGAAGGGTGG 2116  
Qy 792 CTGGAGCGCTGCTGAGAGGATCCATGAAGAGGAGTCCGCGAGTGGTGGCGGCTGATT 851  
Db 2115 CTGGAGCGCTGCTGAGAGGATCCATGAAGAGGAGTCCGCGAGTGGTGGCGGCTGATT 2056  
Qy 852 GATGTGATCGACTGGAAACACCTTCGAATATCTGGGAACTCCCGGAGCGCCAGATCGGC 911  
Db 2055 GATGTGATCGACTGGAAACACCTTCGAATATCTGGGAACTCCCGGAGCGCCAGATCGGC 1996  
Qy 912 GGTTCGACTGAGGCTGTGTGTTTCACTGGGCAACAGTTCCTGAGAGGAGAGATACGG 971  
Db 1995 GGTTCGACTGAGGCTGTGTGTTTCACTGGGCAACAGTTCCTGAGAGGAGAGATACGG 1936  
Qy 972 ATGCAATCCCGCTGCTGATGATCTCAAGTCTCCAAACATGCTGGTGGCTGTTTGTCTG 1031  
Db 1935 ATGCAATCCCGCTGCTGATGATCTCAAGTCTCCAAACATGCTGGTGGCTGTTTGTCTG 1876  
Qy 1032 AGTAAGAAATATTTTGAATATCTGGGCTCTTATGATACAGGAATGAAGTTTGGGAGGA 1091  
Db 1875 AGTAAGAAATATTTGAATATCTGGGCTCTTATGATACAGGAATGAAGTTTGGGAGGA 1816  
Qy 1092 GAAACCTCGAAATTTTCTTTAGGATCTGCGAGTGTGGGCTTCTGGAAACACACCCA 1151





QY 81 ATGTGGGGCGCAGCGCGCGCGCTGCCCGGGGAACTGGCGCGCGCGCGGAGCG 140  
Db 1 ATGTGGGGCGCAGCGCGCGCGCGCTGCCCGGGGAACTGGCGCGCGCGCGGAGCG 60  
QY 141 CTGTGTGTCTCTGGCGCTACTGGCGCTTGGCGGGCTGGGCTCGGTGCTGCGGGCGCAG 200  
Db 61 CTGTGTGTCTCTGGCGCTACTGGCGCTTGGCGGGCTGGGCTCGGTGCTGCGGGCGCAG 120  
QY 201 CTGTGGGCGCGGGCGCGGGCTGCCAGCGCGGACCCCGCGGCACTCCCGCGCGCGCGCGG 260  
Db 121 CTGTGGGCGCGGGCGCGGGCTGCCAGCGCGGACCCCGCGGCACTCCCGCGCGCGCGG 180  
QY 261 CGCGAGCGGTTCATGCGCGCGCGCGCGGTGCGCGGAACTGGCGCGCGCGCGGCGAG 320  
Db 181 CGCGAGCGGTTCATGCGCGCGCGCGCGGTGCGCGGAACTGGCGCGCGCGCGGCGAG 240  
QY 321 CGCGTGGCGGTTCAGCTGCGAGGCGAGGCTGGCGGTGCGAGGAGAGCGTGGCGGTG 380  
Db 241 CGCGTGGCGGTTCAGCTGCGAGGCGAGGCTGGCGGTGCGAGGAGAGCGTGGCGGTG 300  
QY 381 CACAGATTAACTTACCTCAGCGACCGCATCTCACTGCAACCGCGCGCTGCCCGAGCGC 440  
Db 301 CACAGATTAACTTACCTCAGCGACCGCATCTCACTGCAACCGCGCGCTGCCCGAGCGC 360  
QY 441 TGGNACCGCTGTGCAAGAGAGAAATATATATATATATATATATATATATATATATAT 500  
Db 361 TGGNACCGCTGTGCAAGAGAGAAATATATATATATATATATATATATATATATATAT 420  
QY 501 ATCATAGCATTTTATATAGAGCTGTGCAACTCTCTCTCGGACAGTTTACAGTGTCTTT 560  
Db 421 ATCATAGCATTTTATATAGAGCTGTGCAACTCTCTCTCGGACAGTTTACAGTGTCTTT 480  
QY 561 GAGACATCCCCGGATATCTCTAGAGAGAGTATCTCTTGTAGTACTACAGTATAGA 620  
Db 481 GAGACATCCCCGGATATCTCTAGAGAGAGTATCTCTTGTAGTACTACAGTATAGA 540  
QY 621 GAGACATCTGAGAGCGCTTGGCCATAGCTTTTCGGGACTGCCAAGAGTGGCGCTGATC 680  
Db 541 GAGACATCTGAGAGCGCTTGGCCATAGCTTTTCGGGACTGCCAAGAGTGGCGCTGATC 600  
QY 681 CGCGCCAAAGAGAGAGGCGCTGTGCGAGCGCGCGCTGTGGGGGCTGTGGCGGAGG 740  
Db 601 CGCGCCAAAGAGAGAGGCGCTGTGCGAGCGCGCGCTGTGGGGGCTGTGGCGGAGG 660  
QY 741 GCGCATGTTCTGACCTTCTCTGACCTGTCACTGTGAGTGCACGAAGGCTGGCGCGC 800  
Db 661 GCGCATGTTCTGACCTTCTCTGACCTGTCACTGTGAGTGCACGAAGGCTGGCGCGC 720  
QY 801 CTGCTGAGAGGATCCATGAAGAGAGTGGCGAGTGGTGTGCCCGGTGATGTGATC 860  
Db 721 CTGCTGAGAGGATCCATGAAGAGAGTGGCGAGTGGTGTGCCCGGTGATGTGATC 780  
QY 861 GACTTGAACACCTTCGAATACCTGGGAACTCCGGGAGCGCCAGATCGCGGTTCGAC 920  
Db 781 GACTTGAACACCTTCGAATACCTGGGAACTCCGGGAGCGCCAGATCGCGGTTCGAC 840  
QY 921 TGGAGGCTGTGTTCAGTGGGCAACAGTTTCTGAGAGGAGAGGATACGGATGCAATCC 980  
Db 841 TGGAGGCTGTGTTCAGTGGGCAACAGTTTCTGAGAGGAGAGGATACGGATGCAATCC 900  
QY 981 CCGGTGATGTCATCAGGTCTCAACAAATGGGTGGGCTGTGTTGCTGTGAGTAAAGAA 1040  
Db 901 CCGGTGATGTCATCAGGTCTCAACAAATGGGTGGGCTGTGTTGCTGTGAGTAAAGAA 960  
QY 1041 TATTTTGAATATCTGGGTCTTATGATACAGGAATGGAAGTTTGGGAGAGAGAAACCTC 1100  
Db 961 TATTTTGAATATCTGGGTCTTATGATACAGGAATGGAAGTTTGGGAGAGAGAAACCTC 1020  
QY 1101 GAATTTTCTTTAGGATCTGGCAGTGTGGGTCTCTGGAACACACCCATGTTCCTCAT 1160  
Db 1021 GAATTTTCTTTAGGATCTGGCAGTGTGGGTCTCTGGAACACACCCATGTTCCTCAT 1080  
QY 1161 GTTGGGCATGTTTTTCCCAAGAGCTCCCTACTCCCGCAACAGGCTCTGGCCACAGT 1220

Db 1081 GTTGGGCATGTTTTTCCCAAGAGCTCCCTACTCCCGCAACAGGCTCTGGCCACAGT 1140  
QY 1221 GTTTCGTGCGAGCTGAAGTATGGATGAATTTAAAGAGCTCTACTTACCATCGCAACCCC 1280  
Db 1141 GTTTCGTGCGAGCTGAAGTATGGATGAATTTAAAGAGCTCTACTTACCATCGCAACCCC 1200  
QY 1281 CGTGGCCGCTTGAACTTTTGGGATGTGACAGAGAGGAGAGCTCCGGGCAAGCTC 1340  
Db 1201 CGTGGCCGCTTGAACTTTTGGGATGTGACAGAGAGGAGAGCTCCGGGCAAGCTC 1260  
QY 1341 CAGTGTAAAGACTTCAAGTGGTCTTTGGAGACTGTGTATCCAGAACTGATGTGCTGAG 1400  
Db 1261 CAGTGTAAAGACTTCAAGTGGTCTTTGGAGACTGTGTATCCAGAACTGATGTGCTGAG 1320  
QY 1401 GACAGCGCTTGGCTCTTTGGGATGTCTCCAGAACAAAGAGCTAAACAGACTACTGCTTTGAC 1460  
Db 1321 GACAGCGCTTGGCTCTTTGGGATGTCTCCAGAACAAAGAGCTAAACAGACTACTGCTTTGAC 1380  
QY 1461 TATAACCTCTCCGATGAAACCCAGATTGTGGGACACAGGTCACTTGTACCTCTGTCTAT 1520  
Db 1381 TATAACCTCTCCGATGAAACCCAGATTGTGGGACACAGGTCACTTGTACCTCTGTCTAT 1440  
QY 1521 GGGATGGGCGCAGAACTCAGTTTTTTCGAGTACACGTCCAGAACAAAGAAATACGCTATAACCC 1580  
Db 1441 GGGATGGGCGCAGAACTCAGTTTTTTCGAGTACACGTCCAGAACAAAGAAATACGCTATAACCC 1500  
QY 1581 CACACGCTTGAGGCGTGCATTTGCTGTGGAGCAGGAAATGATACCTTATCATGCACTCTC 1640  
Db 1501 CACACGCTTGAGGCGTGCATTTGCTGTGGAGCAGGAAATGATACCTTATCATGCACTCTC 1560  
QY 1641 TCGAAGAACTGCCCGCAGAGAACTCAGAAAGTTTCATCTTCAGGAGGATGGATCTTTATTT 1700  
Db 1561 TCGAAGAACTGCCCGCAGAGAACTCAGAAAGTTTCATCTTCAGGAGGATGGATCTTTATTT 1620  
QY 1701 CACGAACAGTCCAGAAATGTTCAGGCTGCAGGCTGCAGGAAGAGTGCAGTGTTCGTT 1760  
Db 1621 CACGAACAGTCCAGAAATGTTCAGGCTGCAGGCTGCAGGAAGAGTGCAGTGTTCGTT 1680  
QY 1761 CCACTCTTACGAGACTGCCCACTCGGATCATCAGAAATGGTCTTCAAGAGCGCATG 1820  
Db 1681 CCACTCTTACGAGACTGCCCACTCGGATCATCAGAAATGGTCTTCAAGAGCGCATG 1740  
QY 1821 TTATGA 1826  
Db 1741 TTATGA 1746

## RESULT 4

US-10-292-896-57

; Sequence-57 Application US/10292896

; Publication No. US20030186850A1

; GENERAL INFORMATION:

; APPLICANT: HASSAN, Helle

; APPLICANT: REIS, Celso A.

; APPLICANT: BENNETT, Eric P.

; APPLICANT: CLAUSEN, Henrik

; TITLE OF INVENTION: METHODS TO IDENTIFY AGENTS MODULATING FUNCTIONS OF POLYPEPTIDE GAI

; TITLE OF INVENTION: TRANSFERASES, PHARMACEUTICAL COMPOSITIONS COMPRISING SUCH AGENTS

; FILE REFERENCE: 4305/1H154-US3

; CURRENT APPLICATION NUMBER: US/10/292,896

; CURRENT FILING DATE: 2003-03-17

; PRIOR APPLICATION NUMBER: US 60/425,204

; PRIOR FILING DATE: 2002-11-08

; PRIOR APPLICATION NUMBER: PCT/DK01/00328

; PRIOR FILING DATE: 2001-05-10

; PRIOR APPLICATION NUMBER: US 60/203,331

; PRIOR FILING DATE: 2000-05-11

; NUMBER OF SEQ ID NOS: 130

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 57

; LENGTH: 1746





367	TTTCGGGACTGCCCAAGGTGGCCCTGATCCGCGCCAAACAAGAGAGAGGGGCTCGTGGCAG	426
712	CCCGGCTGCTGGGGCGCTCTCGCGCGAGGGGCGATGTTCTTGACCTTCTCGACTGTCACT	771
427	CCCGGCTGCTGGGGCGCTCTCGCGCGARGGGCGATGTTCTGACCTTCTCGACTGTCACT	486
772	GTGAGTCCCAAGAGGTGGCTGAGCGCGCTGCTGCAGAGGATCCATGAAGAGAGTGG	831
487	GTGAGTCCCAAGAGGT - GCTGAGAGCGCTGCTGCAGAGGATCCATGAAGAGAGTGG	545
832	CAGTGGTGTGCCCGGTGATTGATCTGATCGACTGGAAACACCTTCCGAATACCTGGGGAAC	891
546	CAGTGGTGTGCCCGGTGATTGATGTSATCGACTGGAACACCTTCCGAATACCTGGGGAAC	605
892	CCGGGAGCCCCAGATCGGCGGTTTTGCACTGGAGGCTGGTGTTCACGTGGGCACACAGTTC	951
606	CCGGGAGCCCCAGATCGGCGGTTTTGCACTGGAGGCTGGTGTTCACGTGGGCACACAGTTC	665
952	CTGAGAGGGAGAGATACGGATCGAATCCCGCTCGATGTCATCAGGCTTCCCAATATGG	1011
666	CTGAGAGGGAGAGATACGGATCGAATCCCGCTCGATGTCATCAGGCTTCCCAATATGG	725
1012	CTGCTGGGCTGTTTCTGTGAGTGAAGAAATATTTTGAATATCTGGGGTCTTATGATACAG	1071
726	CTGCTGGGCTGTTTCTGTGAGTGAAGAAATATTTTGAATATCTGGGGTCTTATGATACAG	785
1072	GAATGGAAGTTTGGGGAGGAAAAACCTCGAATTTTCTTTAGGATCTGGCAGTGTGGTG	1131
786	GAATGGAAGTTTGGGGAGGAAAAACCTCGAATTTTCTTTAGGATCTGGCAGTGTGGTG	845
1132	GGGTTCTGGAACACACCCATGTTCCCATGTTGGCCATGTTTTCCCACGAAGCTCCCT	1191
846	GGGTTCTGGAACACACCCATGTTCCCATGTTGGCCATGTTTTCCCACGAAGCTCCCT	905
1192	ACTCCCGCAACAGGCTCTGGCCAAACAGTGTTCGTGCACTGGAAGTATGGAATGAAT	1251
906	ACTCCCGCAACAGGCTCTGGCCAAACAGTGTTCGTGCACTGGAAGTATGGAATGAAT	965
1252	TTAAGAGCTCTACTACCATCGCAACCCCGTGCCCGCTTGGAAACCTTTTGGGAGTGTGA	1311
966	TTAAGAGCTCTACTACCATCGCAACCCCGTGCCCGCTTGGAAACCTTTTGGGAGTGTGA	1025
1312	CAGAGAGGAAGCAGCTCCGGGACAAGCTCCAGTGTAAAGACTTCAAGTGGTTCCTGGAGA	1371
1026	CAGAGAGGAAGCAGCTCCGGGACAAGCTCCAGTGTAAAGACTTCAAGTGGTTCCTGGAGA	1085
1372	CTGTGTATCCAGAACTGCATGTGCTGAGGACAGGCTTGGCTTCTTGGGATGCTCCAGA	1431
1086	CTGTGTATCCAGAACTGCATGTGCTGAGGACAGGCTTGGCTTCTTGGGATGCTCCAGA	1145
1432	ACAAAGGACTAACAGACTACTGCTTTTGACTATAACCTCCCGATGAAAAACAGATTTGG	1491
1146	ACAAAGGACTAACAGACTACTGCTTTTGACTATAACCTCCCGATGAAAAACAGATTTGG	1205
1492	GACACCAAGGTCAATCTGTGACCTCTGTGTCATGGATGGGCCAGAAATCA	1537
1206	GACACCAAGGTCAATCTGTGACCTCTGTGTCATGGATGGGCCAGAAATCA	1251

## RESULT 7

RESULT /  
US-10-106-698-598

US-10-106-698-398  
: Sequence 598. Application US/10106698; sequence 398, Application US/10  
; Publication No. US20030109690A1; Publication No: US20  
: GENERAL INFORMATION:

; GENERAL INFORMATION:  
; APPLICANT: Ruben et al.

;  
; AFFILICANI: RUBEN ET AL.  
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide

; TITLE OF INVENTION: CO  
; FILE REFERENCE: PA005P1

FILE REFERENCE: FPO0001  
CURRENT APPLICATION NUMBER: US/10/106,698

; CURRENT AFFILIATION NUMBER: 03/10/100,000  
 ; CURRENT FILING DATE: 2002-03-27

; CURRENT FILING DATE: 2002-03-27  
 ; PRIOR APPLICATION NUMBER: PCT/US00/26524

; PRIOR AFFIDAVIT NUMBER: FCH  
 ; PRIOR FILING DATE: 2000-09-28

; PRIOR FILING DATE: 2000-09-28  
; PRIOR APPLICATION NUMBER: US 60/157,137

; PRIOR FILING\_DATE: 1999-09-29

```

; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 598
; LENGTH: 1517
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (144)..(144)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-106-698-598

```

Query Match	35.5%	Score 1012.2	DB 14	Length 1517	
Best Local Similarity	99.3%	Pred. No. 2.7e-224			
Matches 1008	Conservative	7	Mismatches 0	Indels 0	Gaps 0
QY	1115	GATCTGCCAGTGTGGTGGGTCTCGGAAACACACACCACATGTTCCCATGTTGGCCATGTTTT	1174		
DB	479	GAITCTGCGAGTGTGGTGGGTCTCGGAAACACACACCACATGTTCCCATGTTGGCCATGTTTT	538		
QY	1175	CCCCAAGCAAGCTCCCTTACTCCCGCAACAAAGGCTCTGGCCAAACAGTGTTCGTGCAAGCTGA	1234		
DB	539	CCCCAAGCAAGCTCCCTTACTCCCGCAACAAAGGCTCTGGCCAAACAGTGTTCGTGCAAGCTGA	598		
QY	1235	AGTATGATCGATGAATTTAAAGAGCTCTACTACCATCGCAACCCCGTCCCGCTTGGGA	1294		
DB	599	AGTATGATCGATGAATTTAAAGAGCTCTACTACCATCGCAACCCCGTCCCGCTTGGGA	658		
QY	1295	ACCTTTTGGGGATGTGACAGAGAGGAAGCAGCTCCGGGACAAAGCTCCAGTGTAAAGACTT	1354		
DB	659	ACCTTTTGGGGATGTGACAGAGAGGAAGCAGCTCCGGGACAAAGCTCCAGTGTAAAGACTT	718		
QY	1355	CAAGTGGTCTTTGGAGACTGTGTATCCAGAACTGCAATGCTGCTGAGACAGGCGCTGCGTT	1414		
DB	719	CAAGTGGTCTTTGGAGACTGTGTATCCAGAACTGCAATGCTGCTGAGACAGGCGCTGCGTT	778		
QY	1415	CTTCGGGATGCTCCAGAAACAAAGACTTAAACAGACTACTGCTTTTGACTATTAACCCCTCCCGA	1474		
DB	779	CTTCGGGATGCTCCAGAAACAAAGACTTAAACAGACTACTGCTTTTGACTATTAACCCCTCCCGA	838		
QY	1475	TGAAAAACAGATTGTGGGACACACAGGTCTATTCTGTACTCTGTGATGGGATGGGCCAGAA	1534		
DB	839	TGAAAAACAGATTGTGGGACACACAGGTCTATTCTGTACTCTGTGATGGGATGGGCCAGAA	898		
QY	1535	TCAGTTTTTCGAGTACACGCTCCAGAAAGAAATAACGCTATAACACCCACACAGCCTGAGGG	1594		
DB	899	TCAGTTTTTCGAGTACACGCTCCAGAAAGAAATAACGCTATAACACCCACACAGCCTGAGGG	958		
QY	1595	CTGCATTGCTGTGGAAGCAGGAATGGATACCCCTTATCATGCATCTCTGCGAAGAAATGTC	1654		
DB	959	CTGCATTGCTGTGGAAGCAGGAATGGATACCCCTTATCATGCATCTCTGCGAAGAAATGTC	1018		
QY	1655	CCCAGAGNATCAGAAAGTTTCATCTTGCAGGAGGATGGATCTTTATTTACCGAACAGTCCAA	1714		
DB	1019	CCCAGAGNATCAGAAAGTTTCATCTTGCAGGAGGATGGATCTTTATTTACCGAACAGTCCAA	1078		
QY	1715	GAATATGCTCCAGGCTCGAGGAAGAGTGCAGTGACAGTTTCGTTCCACTCTTACGAGA	1774		
DB	1079	GAATATGCTCCAGGCTCGAGGAAGAGTGCAGTGACAGTTTCGTTCCACTCTTACGAGA	1138		
QY	1775	CTGCACCGAATCTGGGATCATCAGAAATGGTCTTCAAAGAGGCGATGTTATGCAAGCCTCGT	1834		
DB	1139	CTGCACCGAATCTGGGATCATCAGAAATGGTCTTCAAAGAGGCGATGTTATGCAAGCCTCGT	1198		
QY	1835	GTATCAAGGAGCCCATCGAAGGACACTGTGGAGCCAGGACTCTGCCCAAACAAAGACTTAG	1894		
DB	1199	GTATCAAGGAGCCCATCGAAGGACACTGTGGAGCCAGGACTCTGCCCAAACAAAGACTTAG	1258		
QY	1895	CTAAGCAGTGACAGAAACCCACCAAAATCTAGGCTGTGATTTGCTTTGAAGAGGCAATCATTT	1954		
DB	1259	CTAAGCAGTGACAGAAACCCACCAAAATCTAGGCTGTGATTTGCTTTGAAGAGGCAATCATTT	1318		

QY 1955 TTGCCATTCTGGAAGTTGTTGGATTAGTAAATAATGGAATAAGCTTTGTTACTTTATT 2014  
Db 1319 TTGCCATTCTGGAAGTTGTTGGATTAGTAAATAATGGAATAAGCTTTGTTACTTTATT 1378  
QY 2015 TTGAGAACTTTTAAATGTTCCAAATACCTCTATTTTCAAAGGGTAAATCGTAAGATGTTA 2074  
Db 1379 TTGAGAACTTTTAAATGTTCCAAATACCTCTATTTTCAAAGGGTAAATCGTAAGATGTTA 1438  
QY 2075 ACCCTTGGTATTAGAAAAATTAACCTTATTAATTTTCTATCAARAWMAWA 2129  
Db 1439 ACCCTTGGTATTAGAAAAATTAACCTTATTAATTTTCTAWMAAAAAAAA 1493

RESULT 8  
US-09-895-298-44  
; Sequence 44, Application US/09895298  
; Publication No. US20030078405A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: 47 Human Secreted Proteins  
; FILE REFERENCE: P2035P1  
; CURRENT APPLICATION NUMBER: US/09/895,298  
; CURRENT FILING DATE: 2001-07-02  
; PRIOR APPLICATION NUMBER: 09/591,16  
; PRIOR FILING DATE: 2000-06-09  
; PRIOR APPLICATION NUMBER: PCT/US99/29950  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: 60/113,006  
; PRIOR FILING DATE: 1998-12-18  
; PRIOR APPLICATION NUMBER: 60/112,809  
; PRIOR FILING DATE: 1998-12-17  
; NUMBER OF SEQ ID NOS: 231  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 44  
; LENGTH: 1517  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (144)  
; OTHER INFORMATION: n equals a,t,g, or c  
US-09-895-298-44

Query Match 35.5%; Score 1011.8; DB 11; Length 1517;

Best Local Similarity 99.2%; Pred. No. 3.3e-224;

Matches 1007; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

QY 1115 GATCTGGCAGTGTGGTGGGTTCTGGAACACACCCCATGTTCCCATGTTGGCCATGTTTT 1174  
Db 479 GATCTGGCAGTGTGGTGGGTTCTGGAACACACCCCATGTTCCCATGTTGGCCATGTTTT 538  
QY 1175 CCCAAGCAGCTCCCTACTCCCGCAACAGGCTCTGGCCAAACAGTGTTCGTCAGCTGA 1234  
Db 539 CCCAAGCAGCTCCCTACTCCCGCAACAGGCTCTGGCCAAACAGTGTTCGTCAGCTGA 598  
QY 1235 AGTATGGATGGATGAATTTAAAGAGCTCTACTACCATCGCAACCCCGTCCCGCTTGA 1294  
Db 599 AGTATGGATGGATGAATTTAAAGAGCTCTACTACCATCGCAACCCCGTCCCGCTTGA 658  
QY 1295 ACCTTTGGGATGTGACAGAGGAAAGAGCTCCGGGCAAGCTCCAGTGTAAAGACTT 1354  
Db 659 ACCTTTGGGATGTGACAGAGGAAAGAGCTCCGGGCAAGCTCCAGTGTAAAGACTT 718  
QY 1355 CAAGTGTCTTCGAGACTGTGATCCAGAACTGCTGCTCAGACAGGCTCGCTT 1414  
Db 719 CAAGTGTCTTCGAGACTGTGATCCAGAACTGCTGCTCAGACAGGCTCGCTT 778  
QY 1415 CTTTCGGATGCTCCAGAAACAAAGACTTAACAGACTACTGCTTTGACTATACCTCCCGA 1474  
Db 779 CTTTCGGATGCTCCAGAAACAAAGACTTAACAGACTACTGCTTTGACTATACCTCCCGA 838  
QY 1475 TGAACACCAAGTTGTGGGACACCAAGGTCATCTGTACCTCTGTATGGGATGGGCCAGAA 1534

Db 839 TGAACACCAAGTTGTGGGACACCAAGGTCATCTGTACCTCTGTATGGGATGGGCCAGAA 898  
QY 1535 TCAGTTTTTTCGAGTACACGTCCTCCAGAAAGAAATACGCTATAACACCCACGAGCTGAGGG 1594  
Db 899 TCAGTTTTTTCGAGTACACGTCCTCCAGAAAGAAATACGCTATAACACCCACGAGCTGAGGG 958  
QY 1595 CTGCATTGCTGTGGAGCAGGAATGGATACCTTTATCATCATCTCTGTGCGAAGAACTGC 1654  
Db 959 CTGCATTGCTGTGGAGCAGGAATGGATACCTTTATCATCATCTCTGTGCGAAGAACTGC 1018  
QY 1655 CCAGAGAAATCAGAACTTCATCTTCAGGAGGATGGATCTTTATTTTCCACCAAGCTCAA 1714  
Db 1019 CCAGAGAAATCAGAACTTCATCTTCAGGAGGATGGATCTTTATTTTCCACCAAGCTCAA 1078  
QY 1715 GAAATGTGTCAGGCTCGAGGAAGGAGTGCAGTGCAGTTTCTGTTCCACTCTTTACGAGA 1774  
Db 1079 GAAATGTGTCAGGCTCGAGGAAGGAGTGCAGTGCAGTTTCTGTTCCACTCTTTACGAGA 1138  
QY 1775 CTGCACAACTCGGATCATCAGAAATGGTTTCTTCAAAGAGCGCATGTTATGAAGCTCGT 1834  
Db 1139 CTGCACAACTCGGATCATCAGAAATGGTTTCTTCAAAGAGCGCATGTTATGAAGCTCGT 1198  
QY 1835 GTATCAAGAGCCCATCGAAGGAGACTGTGGAGCCAGGACTCTGCCCAACAAAGACTTAG 1894  
Db 1199 GTATCAAGAGCCCATCGAAGGAGACTGTGGAGCCAGGACTCTGCCCAACAAAGACTTAG 1258  
QY 1895 CTAAGCAGTGCACAGAACCCACCAAACTTAGCTGCTGCTTTTGAAGAGGCAATCAT 1954  
Db 1259 CTAAGCAGTGCACAGAACCCACCAAACTTAGCTGCTGCTTTTGAAGAGGCAATCAT 1318  
QY 1955 TTGCCATTTGTGAAGTTGTTGGATTTAGTAAAAATGTGAATAAGCTTTGTACTTATT 2014  
Db 1319 TTGCCATTTGTGAAGTTGTTGGATTTAGTAAAAATGTGAATAAGCTTTGTACTTATT 1378  
QY 2015 TTGAGAACTTTTAAATGTTCCAAATACCTCTATTTTCTATCAARAWMAWA 2129  
Db 1379 TTGAGAACTTTTAAATGTTCCAAATACCTCTATTTTCTATCAARAWMAWA 1438  
QY 2075 ACCCTTGGTATTAGAAAAATTAACCTTATTAATTTTCTAWMAAAAAAAA 1493  
Db 1439 ACCCTTGGTATTAGAAAAATTAACCTTATTAATTTTCTAWMAAAAAAAA 1493

## RESULT 9

US-10-198-846-7621

; Sequence 7621, Application US/10198846

; Publication No. US20030099974A1

; GENERAL INFORMATION:

; APPLICANT: Lillie, James

; APPLICANT: Xu, Yongyao

; APPLICANT: Wang, Youzhen

; APPLICANT: Steimann, Kathleen

; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS

; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND

; FILE REFERENCE: MRI-049

; CURRENT APPLICATION NUMBER: US/10/198,846

; PRIOR FILING DATE: 2002-07-18

; PRIOR FILING DATE: 2001-07-18

; NUMBER OF SEQ ID NOS: 14084

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 7621

; LENGTH: 818

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: 2\_3\_16, 529, 689, 691, 709, 713, 720, 726, 739, 741, 742,

; LOCATION: 746, 749, 755, 759, 760, 761, 767, 768, 774, 775, 777, 783,

; LOCATION: 790, 791, 797, 798, 799, 800, 801, 802, 803, 804, 805, 806,

; LOCATION: 811, 814, 815, 816, 817

; OTHER INFORMATION: n = A,T,C or G  
US-10-198-846-7621

Query Match 20.5%; Score 584.6; DB 14; Length 818;  
Best Local Similarity 99.2%; Pred. No. 3.3e-125;  
Matches 587; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
Qy 2243 CAACGTACTTTGCAATTAACGTAAATACCTCAGCTGGGGGTTAAAGTTTCCAGTATA 2302  
Db 96 CGAGGTACTTTGCAATTAACGTAAATACCTCAGCTGGGGGTTAAAGTTTCCAGTATA 155  
Qy 2303 GAGAGACTGTCACCTAGGAACAATGTTGATTTATTCAGGTCAATGAGATCTTCTAGATG 2362  
Db 156 GAGAGACTGTCACCTAGGAACAATGTTGATTTATTCAGGTCAATGAGATCTTCTAGATG 215  
Qy 2363 TATTTTAAAAAGAAATGCTTTTGTGTTATGTTGCTACCAACAGTTAAACATCCCAATATGT 2422  
Db 216 TATTTTAAAAAGAAATGCTTTTGTGTTATGTTGCTACCAACAGTTAAACATCCCAATATGT 275  
Qy 2423 TCATGTCAGCCCAAGAGGACTAAACAAAGCTGAAATCTCAGAGAACAAATTTGCTTTACTA 2482  
Db 276 TCATGTCAGCCCAAGAGGACTAAACAAAGCTGAAATCTCAGAGAACAAATTTGCTTTACTA 335  
Qy 2483 AGCTGAGTCAACTTGAGAGCGAACTTCTAACAAATGCCGCACTGTAGTGTGGCTGGTTCTA 2542  
Db 336 AGCTGAGTCAACTTGAGAGCGAACTTCTAACAAATGCCGCACTGTAGTGTGGCTGGTTCTA 395  
Qy 2543 CCACATGACATTTAAACATGTTTATATCAATTTTAAATTTTATGATGAGTGTGTCAG 2602  
Db 396 CCACATGACATTTAAACATGTTTATATCAATTTTAAATTTTATGATGAGTGTGTCAG 455  
Qy 2603 GGAGAAATGTAATGTTCTATATGAAATCCCTTTTCAAGTTTGTTCATTAATACAGTTA 2662  
Db 456 GGAGAAATGTAATGTTCTATATGAAATCCCTTTTCAAGTTTGTTCATTAATACAGTTA 515  
Qy 2663 TTAATTTAAATCAGGTTAGAGTTTGTGCTGCTGCAACTGCTGTGAAATTTCTCTGAGT 2722  
Db 516 TTAATTTAAATCANGTTAGAGTTTGTGCTGCTGCAACTGCTGTGAAATTTCTCTGAGT 575  
Qy 2723 AATTCGTGTTGTAATGATCCAGACCAACCCCTGAGATTTTGTCAACCTGATTAAGTCA 2782  
Db 576 AATTCGTGTTGTAATGATCCAGACCAACCCCTGAGATTTTGTCAACCTGATTAAGTCA 635  
Qy 2783 ATATGATGATTAAGAGATGTGAGACCAAAAAAAGAAAAAAGAAAAA 2834  
Db 636 ATATGATGATTAAGAGATGTGAGACCAAAAAAAGAAAAAAGAAAAA 687

RESULT 10  
US-09-777-564-414  
; Sequence 414, Application US/09777564  
; Patent No. US20020022591A1  
; GENERAL INFORMATION:  
; APPLICANT: Mannion, Jane  
; APPLICANT: Algate, Paul A.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; FILE REFERENCE: 210121.493  
; CURRENT APPLICATION NUMBER: US/09/777,564  
; CURRENT FILING DATE: 2001-02-05  
; NUMBER OF SEQ ID NOS: 1730  
; SOFTWARE: FastSeq for Window Version 4.0  
; SEQ ID NO 414  
; LENGTH: 544  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(544)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-777-564-414

Query Match 18.5%; Score 526.4; DB 9; Length 544;

Best Local Similarity 99.6%; Pred. No. 8.3e-112;  
Matches 538; Conservative 0; Mismatches 1; Indels 1; Gaps 1;  
Qy 1407 CTGGGCTTCTCGGATGCTCCAGAACAAAGGACTAACAGACTACTGCTTTGACTATAAC 1466  
Db 1 CCTGGCTTCTTCGGGATGCTCCAGAACAAAGGACTAACAGACTACTGCTTTGACTATAAC 60  
Qy 1467 CTCCCGATGAAACACAGATTGTTGGGACACACAGGTCAATCTGTACCTCTGTCTATGGGATG 1526  
Db 61 CTCCCGATGAAACACAGATTGTTGGGACACACAGGTCAATCTGTACCTCTGTCTATGGGATG 120  
Qy 1527 GGCCAGAAATCAGTTTTTTCGAGTACACGTCCTCCAGAAAGAAATACGCTATACACCCACAG 1586  
Db 121 GGCCAGAAATCAGTTTTTTCGAGTACACGTCCTCCAGAAAGAAATACGCTATACACCCACAG 180  
Qy 1587 CTGTGGGCTGCATTCGTTGGGAGCAGGAATGATACCTTATCATGCTCTCTCGAA 1646  
Db 181 CTGTGGGCTGCATTCGTTGGGAGCAGGAATGATACCTTATCATGCTCTCTCGAA 240  
Qy 1647 GAAACTGCCCCAGAGAAATCAGAAATGTTTCATCTTCAGAGAGATGATCTTTTATTTTCAGAA 1706  
Db 241 GAAACTGCCCCAGAGAAATCAGAAATGTTTCATCTTCAGAGAGATGATCTTTTATTTTCAGAA 300  
Qy 1707 CAGTCCAAAGAAATGTTTCAGGCTGCGAGGAGAGTTCGAGTGCACAGTTTCGTTCCACTC 1766  
Db 301 CAGTCCAAAGAAATGTTTCAGGCTGCGAGGAGAGTTCGAGTGCACAGTTTCGTTCCACTC 360  
Qy 1767 TTACGAGACTGCACCAACTCGGATCATCAGAAATGTTTCTTCAAGAGCGCATGTTATGA 1826  
Db 361 TTACGAGACTGCACCAACTCGGATCATCAGAAATGTTTCTTCAAGAGCGCATGTTATGA 420  
Qy 1827 AGCTCGTGTATCAAGAGCGCCATCAAGAGGAGAGTGTGGAGCCAGGACTCTGCCCAACAA 1886  
Db 421 AGCTCGTGTATCAAGAGCGCCATCAAGAGGAGAGTGTGGAGCCAGGACTCTGCCCAACAA 480  
Qy 1887 AGACTTAGCTAAGCAGTGCAGAACCCACCAAAACTAGGCTTGCACTTCTTGAAGAG 1945  
Db 481 AGACTTAGCTAAGCAGTGCAGAACCCACCAAAACTAGGCTTGCACTTCTTGAAGAG 540

RESULT 11  
US-10-015-219-414  
; Sequence 414, Application US/10015219  
; Publication No. US20030008299A1  
; GENERAL INFORMATION:  
; APPLICANT: Algate, Paul A.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; FILE REFERENCE: 210121.493C1  
; CURRENT APPLICATION NUMBER: US/10/015,219  
; CURRENT FILING DATE: 2002-03-02  
; NUMBER OF SEQ ID NOS: 1739  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 414  
; LENGTH: 544  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 544  
; OTHER INFORMATION: n = A,T,C or G  
US-10-015-219-414  
Query Match 18.5%; Score 526.4; DB 14; Length 544;  
Best Local Similarity 99.6%; Pred. No. 8.3e-112;  
Matches 538; Conservative 0; Mismatches 1; Indels 1; Gaps 1;  
Qy 1407 CCTGGCTTCTTCGGGATGCTCCAGAACAAAGGACTAACAGACTACTGCTTTGACTATAAC 1466  
Db 1 CCTGGCTTCTTCGGGATGCTCCAGAACAAAGGACTAACAGACTACTGCTTTGACTATAAC 60  
Qy 1467 CCTCCGATGAAACACAGATTGTTGGGACACACAGGTCAATCTGTACCTCTGTCTATGGGATG 1526

Db 61 CCTCCGATGAAACACAGATTGTGGGACACAGGTCACTCTGTACCTCTGTCTGTCTGGATG 120  
Qy 1527 GGCAGAAATCAGTTTTTTCAGTACACGTCCAGAAAGAAATACGCTATAACACCCACAG 1586  
Db 121 GGCAGAAATCAGTTTTTTCAGTACACGTCCAGAAAGAAATACGCTATAACACCCACAG 180  
Qy 1587 CCTGAGGCTGCATTGCTGTGGAGCAGGAGTATGATACCTTATCATGATCTCTCTGGAA 1646  
Db 181 CCTGAGGCTGCATTGCTGTGGAGCAGGAGTATGATACCTTATCATGATCTCTCTGGAA 240  
Qy 1647 GAAATCTGCCAGAGAAATCAGAAATTCATCTTGCAGGAGGATGATCTTTATTTACAGAA 1706  
Db 241 GAAATCTGCCAGAGAAATCAGAAATTCATCTTGCAGGAGGATGATCTTTATTTACAGAA 300  
Qy 1707 CAGTCCAAAGAAATGTGTCCAGGCTGCGAGGAGGATGATGATGATGATGATGATGATG 1766  
Db 301 CAGTCCAAAGAAATGTGTCCAGGCTGCGAGGAGGATGATGATGATGATGATGATGATG 360  
Qy 1767 TTACGAGACTGCACCACTCGGATCATCAGAAATGTTCTTCAAGAGCGCATGTTATGA 1826  
Db 361 TTACGAGACTGCACCACTCGGATCATCAGAAATGTTCTTCAAGAGCGCATGTTATGA 420  
Qy 1827 AGCTCTGTATCAAGAGCCCATCGAAGGAGACTGTGGAGCCAGGACTCTGCCCAACAA 1886  
Db 421 AGCTCTGTATCAAGAGCCCATCGAAGGAGACTGTGGAGCCAGGACTCTGCCCAACAA 480  
Qy 1887 AGACTTACGTAAGCAGTACACAGAACCCACCAAACTAGGC-TGCATTGCTTTGAAGAG 1945  
Db 481 AGACTTACGTAAGCAGTACACAGAACCCACCAAACTAGGCTTGCTTTGAAGAG 540

## RESULT 12

US-10-292-896-121  
; Sequence 121, Application US/10292896  
; Patent No. US20030186850A1  
; GENERAL INFORMATION:  
; APPLICANT: HASSAN, Helle  
; APPLICANT: REIS, Celso A.  
; APPLICANT: BENNETT, Eric P.  
; APPLICANT: CLAUSEN, Henrik  
; TITLE OF INVENTION: METHODS TO IDENTIFY AGENTS MODULATING FUNCTIONS OF POLYPEPTIDE GR  
; FILE REFERENCE: 4305/1H154-US3  
; CURRENT FILING DATE: 2003-03-17  
; PRIOR FILING DATE: 2003-03-17  
; PRIOR FILING DATE: 2002-11-08  
; PRIOR FILING DATE: 2002-11-08  
; PRIOR FILING DATE: 2001-05-10  
; PRIOR FILING DATE: 2000-05-11  
; NUMBER OF SEQ ID NOS: 130  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 121  
; LENGTH: 525  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-292-896-121

Query Match 18.4%; Score 524; DB 12; Length 525;  
Best Local Similarity 100.0%; Pred. No. 2.9e-111;  
Matches 524; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1303 GGGATGTGACAGAGAGGAGGAGCTCCGGGACAGCTCCAGGTGTAAGACTTCAAGTGGT 1362  
Db 2 GGGATGTGACAGAGAGGAGGAGCTCCGGGACAGCTCCAGGTGTAAGACTTCAAGTGGT 61  
Qy 1363 TCTTGGAGACTGTGTATCCAGAACTGCGCTGAGGAGCAGGCTGGCTTCTTCGGGA 1422  
Db 62 TCTTGGAGACTGTGTATCCAGAACTGCGCTGAGGAGCAGGCTGGCTTCTTCGGGA 121  
Qy 1423 TGCTCCAGAAACAAAGGACTAACAGACTACTGCTTTTGACTATACCTCCCGATGAAAC 1482

Db 122 TGCTCCAGAAACAAAGGACTAACAGACTACTGCTTTGACTATAACCTCCCGATGAAAC 181  
Qy 1483 AGATTGTGGACACACAGGTCACTTCTGTACCTCTGTATGGATGGGCGAGAATCAGTTTT 1542  
Db 182 AGATTGTGGACACACAGGTCACTTCTGTACCTCTGTATGGATGGGCGAGAATCAGTTTT 241  
Qy 1543 TCGAGTACACGTCCCAAGAAATAACGCTATATAACCCACAGCTTGCAGGCTGCATTG 1602  
Db 242 TCGAGTACACGTCCCAAGAAATAACGCTATATAACCCACAGCTTGCAGGCTGCATTG 301  
Qy 1603 CTCTGGAGCAGGAGTATGATACCTTATCATGATCTCTCGGAGAACTGCCCCACAGA 1662  
Db 302 CTCTGGAGCAGGAGTATGATACCTTATCATGATCTCTCGGAGAACTGCCCCACAGA 361  
Qy 1663 ATCAGAAATTCATCTTGCAGGAGGATGATCTTATTTACGAACTGCTTACGAGCTGCACCA 1722  
Db 362 ATCAGAAATTCATCTTGCAGGAGGATGATCTTATTTACGAACTGCTTACGAGCTGCACCA 421  
Qy 1723 TCCAGGCTGCGAGAAAGGAGTTCGAGTGACAGTTTCGTTCCACTCTTACGAGCTGCACCA 1782  
Db 422 TCCAGGCTGCGAGAAAGGAGTTCGAGTGACAGTTTCGTTCCACTCTTACGAGCTGCACCA 481  
Qy 1783 ACTCGGATCATCAGAAATGCTTCTTCAAGAGCGCATGTTATGA 1826  
Db 482 ACTCGGATCATCAGAAATGCTTCTTCAAGAGCGCATGTTATGA 525

## RESULT 13

US-09-777-564-172  
; Sequence 172, Application US/09777564  
; Patent No. US20020022591A1  
; GENERAL INFORMATION:  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Marnion, Jane  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; FILE REFERENCE: 210121.493  
; CURRENT FILING DATE: 2001-02-05  
; NUMBER OF SEQ ID NOS: 1730  
; SOFTWARE: FastSeq for Window Version 4.0  
; SEQ ID NO 172  
; LENGTH: 632  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(632)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-777-564-172

Query Match 17.4%; Score 496.4; DB 9; Length 632;  
Best Local Similarity 90.6%; Pred. No. 8.4e-105;  
Matches 567; Conservative 0; Mismatches 53; Indels 6; Gaps 4;

Qy 1407 CCTGGCTTCTTCGGGATGCTCCAGAACAAAGGACTAACAGACTACTCTTTGACTATAAC 1466  
Db 1 CCTGGCTTCTTCGGGATGCTCCAGAACAAAGGACTAACAGACTACTCTTTGACTATAAC 60  
Qy 1467 CCTCCCGATGAAACACAGATTGTGGACACACAGGTCACTCTGTACCTCTGTCTGGATG 1526  
Db 61 CCTCCCGATGAAACACAGATTGTGGACACACAGGTCACTCTGTACCTCTGTCTGGATG 120  
Qy 1527 GGCAGAAATC-AGTTTTTCAGTACAGTCCCGAGAAAGAAATACGCTATTAACCCACCA 1585  
Db 121 GGCAGAAATCAAGTTTTTTCAGTACACTTCCCGAGAAAGAAATACGCTATTAACCCACCA 180  
Qy 1586 GCCTGAGGCTGCTGCTGTGGAAGCAGGAATGATACCTTATCATGCTCTCTCGGA 1645  
Db 181 GCCTGAGGCTGCTGCTGTGGAAGCAGGAATGATACCTTATCATGCTCTCTCGGA 240  
Qy 1646 AGAACTGCCCCAGAGAAATCAGAAATTCATCTTGCAGGAGGATGATCTTTATTTACGA 1705



Db 241 AGAACTGCCCCAGAGAAATCAGAAATTCATCTTGCAGGAGATGGATCTTTATTTCACGA 300  
Qy 1706 ACAGTCCAGAAATGTGTCCAGGCTCGGAGGAGGAGTCCAGTGCACATTTTCGTTCCACT 1765  
Db 301 ACAGNCCAAGAAATGTGTCCAGGCTCGGAGGAGGAGTCCAGTGCACATTTTCGTTCCACT 360  
Qy 1766 CTTACGAGACTGCACCAACTCGGATCATCAGAAATGGTCTTCAAAGAGCGCATGTTATG 1825  
Db 361 CTTACGAGACTGCACCAACTCGGATCATCAGAAATGGTCTTCAAAGAGCGCATGTTATG 420  
Qy 1826 AAGCCTCGTGTATCAAGGAGCCCATCGAAGGAGACTGTGGAGCCAGGACTCTGCCCAACA 1885  
Db 421 AAGCCTCGTGTATCAANGAGCCCATCGAAGGAGACTGTGGAGCCAGGACTCTGCCCAACA 479  
Qy 1886 AAGACTTAGTAAAGAGTACAGAAACCCACCAAACTAGGCT--GCATTTGCTTTGAAG 1943  
Db 480 AAGACTTCTAAACAGNGACCCAGAAACCCACCAAACTANGGTTGTATTNCTTTTGAAG 539  
Qy 1944 AGGCAATCATTTTGCATTTGTGAAAGTGTGTTGGATTAGT--AAAAATGTGAATAAG 2001  
Db 540 AAGCAATCATTTTGTGAAAGTGTGTTGGATTAGT--AAAAATGTGAATAAG 599  
Qy 2002 CTTTGTACTTATTGTGAACTTTT 2027  
Db 600 CTTTGGAGCTTTTGTGAAACTTT 625

RESULT 14  
US-09-777-564-1423  
; Sequence 1423, Application US/09777564  
; Patent No. US2002022591A1  
; GENERAL INFORMATION:  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Mannion, Jane  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; FILE REFERENCE: 210121.493  
; CURRENT APPLICATION NUMBER: US/09/777,564  
; CURRENT FILING DATE: 2001-02-05  
; NUMBER OF SEQ ID NOS: 1730  
; SOFTWARE: FastSEQ for Window Version 4.0  
; SEQ ID NO 1423  
; LENGTH: 632  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)...(632)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-777-564-1423

Query Match 17.4%; Score 496.4; DB 9; Length 632;  
Best Local Similarity 90.6%; Pred. No. 8.4e-105;  
Matches 567; Conservative 0; Mismatches 53; Indels 6; Gaps 4;  
Qy 1407 CCTGGCTTCTTCGGGATGCTCCAGAACAAAGGACTAAACAGACTACTGCTTTGACTATAAC 1466  
Db 1 CCTGGCTTCTTCGGGATGCTCCAGAACAAAGGACTAAACAGACTACTGCTTTGACTATAAC 60  
Qy 1467 CCTCCGATGAAACCCAGATTTGGGACACAGGTCATTCTGTACCTCTGTCATGGGATG 1526  
Db 61 CCTCCGATGAAACCCAGATTTGGGACACAGGTCATTCTGTACCTCTGTCATGGGATG 120  
Qy 1527 GGCCAGAAATC-AGTTTTTCGAGTACACGTCCTCCAGAAAGAAATACGCTATAACCCACCA 1585  
Db 121 GGCCAGAAATC-AGTTTTTCGAGTACACGTCCTCCAGAAAGAAATACGCTATAACCCACCA 180  
Qy 1586 GCCTGAGGCTGCATTTCTGTGGAGGAGGATGATACCCCTTATCGATCTCTGCGA 1645  
Db 181 GCCTGAGGCTGCATTTCTGTGGAGGAGGATGATACCCCTTATCGATCTCTGCGA 240  
Qy 1646 AGAACTGCCCCAGAGAAATCAGAAATTCATCTTGCAGGAGGATGGATCTTTATTTCACGA 1705

Db 241 AGAACTGCCCCAGAGAAATCAGAAATTCATCTTGCAGGAGATGGATCTTTATTTCACGA 300  
Qy 1706 ACAGTCCAGAAATGTGTCCAGGCTCGGAGGAGGAGTCCAGTGCACATTTTCGTTCCACT 1765  
Db 301 ACAGNCCAAGAAATGTGTCCAGGCTCGGAGGAGGAGTCCAGTGCACATTTTCGTTCCACT 360  
Qy 1766 CTTACGAGACTGCACCAACTCGGATCATCAGAAATGGTCTTCAAAGAGCGCATGTTATG 1825  
Db 361 CTTACGAGACTGCACCAACTCGGATCATCAGAAATGGTCTTCAAAGAGCGCATGTTATG 420  
Qy 1826 AAGCCTCGTGTATCAAGGAGCCCATCGAAGGAGACTGTGGAGCCAGGACTCTGCCCAACA 1885  
Db 421 AAGCCTCGTGTATCAANGAGCCCATCGAAGGAGACTGTGGAGCCAGGACTCTGCCCAACA 479  
Qy 1886 AAGACTTAGTAAAGAGTACAGAAACCCACCAAACTAGGCT--GCATTTGCTTTGAAG 1943  
Db 480 AAGACTTCTAAACAGNGACCCAGAAACCCACCAAACTANGGTTGTATTNCTTTTGAAG 539  
Qy 1944 AGGCAATCATTTTGCATTTGTGAAAGTGTGTTGGATTAGT--AAAAATGTGAATAAG 2001  
Db 540 AAGCAATCATTTTGTGAAAGTGTGTTGGATTAGT--AAAAATGTGAATAAG 599  
Qy 2002 CTTTGTACTTATTGTGAACTTTT 2027  
Db 600 CTTTGGAGCTTTTGTGAAACTTT 625

RESULT 15  
US-10-015-219-172  
; Sequence 172, Application US/10015219  
; Publication No. US20030008299A1  
; GENERAL INFORMATION:  
; APPLICANT: Algate, Paul A.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; FILE REFERENCE: 210121.493C1  
; CURRENT APPLICATION NUMBER: US/10/015,219  
; CURRENT FILING DATE: 2002-03-02  
; NUMBER OF SEQ ID NOS: 1739  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 172  
; LENGTH: 632  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 95, 305, 338, 340, 437, 487, 496, 513, 520, 530, 552, 604  
; OTHER INFORMATION: n = A,T,C or G  
US-10-015-219-172

Query Match 17.4%; Score 496.4; DB 14; Length 632;  
Best Local Similarity 90.6%; Pred. No. 8.4e-105;  
Matches 567; Conservative 0; Mismatches 53; Indels 6; Gaps 4;  
Qy 1407 CCTGGCTTCTTCGGGATGCTCCAGAACAAAGGACTAAACAGACTACTGCTTTGACTATAAC 1466  
Db 1 CCTGGCTTCTTCGGGATGCTCCAGAACAAAGGACTAAACAGACTACTGCTTTGACTATAAC 60  
Qy 1467 CCTCCGATGAAACCCAGATTTGGGACACAGGTCATTCTGTACCTCTGTCATGGGATG 1526  
Db 61 CCTCCGATGAAACCCAGATTTGGGACACAGGTCATTCTGTACCTCTGTCATGGGATG 120  
Qy 1527 GGCCAGAAATC-AGTTTTTCGAGTACACGTCCTCCAGAAAGAAATACGCTATAACCCACCA 1585  
Db 121 GGCCAGAAATC-AGTTTTTCGAGTACACGTCCTCCAGAAAGAAATACGCTATAACCCACCA 180  
Qy 1586 GCCTGAGGCTGCATTTCTGTGGAGGAGGATGATACCCCTTATCGATCTCTGCGA 1645  
Db 181 GCCTGAGGCTGCATTTCTGTGGAGGAGGATGATACCCCTTATCGATCTCTGCGA 240  
Qy 1646 AGAACTGCCCCAGAGAAATCAGAAATTCATCTTGCAGGAGGATGGATCTTTATTTCACGA 1705



Db 241 AGAACTGCCCCAGAGATCAGAAGTTTCATCTTGCAGAGGATGGATCTTTATTTCACGA 300  
Qy 1706 ACAGTCCAGAAATGTGTCCAGGTGCGAGAGAGAGTCGAGTCACAGATTTTCGTCCACT 1765  
Db 301 ACAGNCCAAGAAATGTGTCCAGGTGCGAGAGAGAGTCGAGTCACAGATTTTCGTCCACT 360  
Qy 1766 CTTACGAGACTGCACCACTCGGATCATCAGAAATGGTCTTCAAGAGCGCATGTTATG 1825  
Db 361 CTTACGAGACTGCACCACTCGGATCATCAGAAATGGTCTTCAAGAGCGCATGTTATG 420  
Qy 1826 AAGCTCTGTGTATCAAGGAGCCCATCGAAGGAGACTGTGGAGCCGAGACTCTGCCCAACA 1885  
Db 421 AAGCTCTGTGTATCAANGAGCCCATCGAAGGAGACTGTGGAGCCGAGACTCTGCCCAACA 479  
Qy 1886 AAGACTTAGCTAAGCAGTGACACGAAACCCACCAAACTAGGCT--GCATTTGCTTTGAAG 1943  
Db 480 AAGACTTNCCTAAACAAGNGACCAGAAACCCACCAAACTANGGTTGTATTNCTTTTGAAG 539  
Qy 1944 AGGCAATCATTTTCCCATTTTGTGAAGTTGTGTTGGATTTAGT--AAAAATGTGAATAAG 2001  
Db 540 AAGCAATCATTTNGCCTTTTGTGAAGTTGTGTTGGATTTAAATTAAGGGGGAATAAA 599  
Qy 2002 CTTTGTACTTATTTTGAGAACTTTTT 2027  
Db 600 CTTTNGGACTTTTTTTTGAAAACTTT 625

Search completed: December 8, 2003, 01:56:02  
Job time : 873.526 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 7, 2003, 12:46:13 ; Search time 5923.86 Seconds  
(without alignments)  
11693.015 Million cell updates/sec

Title: US-10-074-527-1

Perfect score: 2850

Sequence: 1 ccggctgggtaccactataa.....aaaaaaaaaaaaaaaaaaaaa 2850

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.\*

1: em\_estba.\*

2: em\_esthum.\*

3: em\_estin.\*

4: em\_estmu.\*

5: em\_estov.\*

6: em\_estpl.\*

7: em\_estro.\*

8: em\_estc.\*

9: gb\_est1.\*

10: gb\_est2.\*

11: gb\_est3.\*

12: gb\_est4.\*

13: gb\_est5.\*

14: gb\_est6.\*

15: em\_estfun.\*

16: em\_estom.\*

17: em\_gss\_hum.\*

18: em\_gss\_inv.\*

19: em\_gss\_pln.\*

20: em\_gss\_vrt.\*

21: em\_gss\_fun.\*

22: em\_gss\_mam.\*

23: em\_gss\_mus.\*

24: em\_gss\_pro.\*

25: em\_gss\_rod.\*

26: em\_gss\_phg.\*

27: em\_gss\_vrl.\*

28: gb\_gss1.\*

29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1303.8	45.7	2217	11 AK033638	AK033638 Mus muscu
2	1297.8	45.5	2074	11 AK042133	AK042133 Mus muscu
3	1257	44.1	2034	11 AK082014	AK082014 Mus muscu
4	1061	37.2	1781	11 BC025639	BC025639 Mus muscu

5	992	34.8	2064	11	BC024475	BC024475 Mus muscu
6	943.8	33.1	1752	11	AK043977	AK043977 Mus muscu
7	868	30.5	909	14	CD251293	CD251293 AGENCOURT
8	738.2	25.9	745	14	CB243715	CB243715 UI-CF-FNO
9	723	25.4	745	12	BM978800	BM978800 UI-CF-DUI
10	718.6	25.2	782	13	BUS84444	BUS84444 AGENCOURT
11	705.4	24.8	711	12	BM977212	BM977212 UI-CF-DUI
12	699.8	24.6	705	12	BM970195	BM970195 UI-CF-ECI
13	666.6	23.4	741	10	AW969574	AW969574 EST381651
14	660.8	23.2	676	10	BE677813	BE677813 7f63d08.x
15	646	22.7	1201	13	BX343372	BX343372 BX343372
16	626	22.0	932	10	EG497683	EG497683 601859717
17	610.8	21.4	967	10	BG167520	BG167520 602342726
18	599.4	21.0	868	10	BE867781	BE867781 601443742
19	580	20.4	1201	13	BX343371	BX343371 BX343371
20	577.4	20.3	792	10	BG687328	BG687328 602639175
21	568.6	20.0	656	10	BG499007	BG499007 602544670
22	566.4	19.9	632	14	CB165548	CB165548 BQ603160
23	557	19.5	559	10	BE218607	BE218607 hv42c04.x
24	548.2	19.2	561	9	AI800923	AI800923 wg14f02.x
25	541.6	19.0	770	14	CD299534	CD299534 AGENCOURT
26	538.6	18.9	708	12	BG777267	BG777267 602664502
27	530.8	18.6	790	10	BF102783	BF102783 601646715
28	524.8	18.4	536	10	BE673969	BE673969 7872b07.x
29	520.8	18.3	524	9	AI630798	AI630798 tx53d05.x
30	520.8	18.3	2607	11	AK033494	AK033494 Mus muscu
31	519.8	18.2	526	9	AI690020	AI690020 tx32c07.x
32	514	18.0	522	9	AI566491	AI566491 tx50e05.x
33	508	17.8	661	12	BI154821	BI154821 602902801
34	505	17.7	518	10	BE502927	BE502927 h281c04.x
35	503.2	17.7	721	12	BG868185	BG868185 602787291
36	491.8	17.3	495	10	BF197099	BF197099 7182b07.x
37	482	16.9	818	10	BF246502	BF246502 601854747
38	480.6	16.9	511	9	AI863865	AI863865 wj55b07.x
39	479	16.8	690	10	BF031122	BF031122 601558805
40	472.4	16.6	477	9	AW612896	AW612896 nh37h04.x
41	465.8	16.3	549	13	BUE79643	BUE79643 UI-CF-ECI
42	457	16.0	457	10	BF222363	BF222363 7p52a04.x
43	456.4	16.0	506	10	BE940623	BE940623 RC3-UT006
44	456.2	16.0	461	9	AI223049	AI223049 qg70b01.x
45	450.8	15.8	496	9	AA429394	AA429394 zw32h12.s

#### ALIGNMENTS

RESULT 1

AK033638

LOCUS

DEFINITION

AK033638

Mus musculus adult male cecum cDNA, RIKEN full-length enriched library, clone:9130206E10 product:hypothetical Glycosyl transferase, family 2 containing protein, full insert sequence.

AK033638

AK033638.1 GI:26329324

HTC; CAP trapper.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning

Meth. Enzymol. 303, 19-44 (1999)

99279253

10349636

2

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

Genome Res. 10 (10), 1617-1630 (2000)

20499374

PUBMED

11042159



QY 542 GACAGTTTACAGTCTCTTGAGACATCCCGGATATCTCTGTAGAGAGTGTCTTGT 601  
Db 468 GACAGTTTACAGTCTCTTGAGACATCCCGGATATCTCTGTAGAGAGTGTCTTGT 527  
QY 602 AGATGACTACAGTCTAGAGACACCTTGAAGAGCGCTTGGCCAAATGAGCTTTCGGGACT 661  
Db 528 AGATGACTACAGTCTAGAGACACCTTGAAGAGCGCTTGGCCAAATGAGCTTTCAGACT 587  
QY 662 GCCCAAGGTGGCTGATCGCGCCAAACAAGAGAGAGGGCTTGGTGGAGCCCGGCTGCT 721  
Db 588 CCCCAAGGTGGCTGATCGCGCCAAACAAGAGAGAGGGCTTGGTGGAGCCCGGCTGCT 647  
QY 722 GGGGGCTCTGCGGAGGGGCGATGTTCTGACCTTCTGACACTGCTGAGTGTGCGCA 781  
Db 648 GGGAGCTCTGCGGAGGGGCGATGTTCTGACCTTCTGACACTGCTGAGTGTGCGCA 707  
QY 782 CGAAGGGTGGCTGAGCGGCTGCTGAGAGGATCCATGAAGAGGAGTGGGAGTGTG 841  
Db 708 TGAGGGTGGCTGAGCGGCTGCTGAGAGGATCCATGAAGAGGAGTGGGAGTGTG 767  
QY 842 CCCGGTGAATGATGATCGATCGAACAACCTTCCGAATACCTGGGGAACCTCCGGGAGCC 901  
Db 768 CCCTGTTATCGATGATGATCGAACAACCTTCCGAATACCTGGGGAACCTCCGGGAGCC 827  
QY 902 CCAGATCGGCGGTTGCACTGGAGGCTGGTGTTCAGCTGGGACACAGTCTCTGAGAGGA 961  
Db 828 GCAGATGGGCGGTTGCACTGGGCGCTGGTATTCACGTGGGACACAGTCTCTGAGAGGA 887  
QY 962 GAGGATACGGATCAATCCCGCTGATGTCATCAGGTCCTCAACAATGGCTGGGCT 1021  
Db 888 CGGCGAGTCAATGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCT 947  
QY 1022 GTTGTCTGTGAGTGAAGAAATATTTGAAATATCTGGGCTCTTATGATACAGGAATGGAAGT 1081  
Db 948 GTTGTCTGTGAGTGAAGAAATATTTGAAATATCTGGGCTCTTATGATACAGGAATGGAAGT 1007  
QY 1082 TTGGGAGGAGAAACCTCGAATTTTCTTTAGATCTGGAGTGTGGGAGTGTGGGAGTGTGGG 1141  
Db 1008 CTGGGAGGAGAAACCTCGAATTTTCTTTAGATCTGGAGTGTGGGAGTGTGGGAGTGTGGG 1067  
QY 1142 AACACACCCATGTTCCCATGTTGGCCATGTTTCCCAAGCAAGCTCCCTACTCCCGCAA 1201  
Db 1068 AACGACCCGCTGCTCCCATGTTGGCCAGCTTCCCTTAAGCAAGCTCCCTACTCCCGTAG 1127  
QY 1202 CAAGGCTCTGGCCAAACAGTGTCCGAGCTGCAGAGTGTGGATGATGATGATGATGATGATGAT 1261  
Db 1128 CAAGGCTCTGGCCAAACAGTGTCCGAGCTGCAGAGTGTGGATGATGATGATGATGATGATGAT 1187  
QY 1262 CTACTACATCGCAACCCCGTGGCCGCTGGGAGTGTGGGAGTGTGGGAGTGTGGGAGTGTGGG 1321  
Db 1188 CTACTACATCGCAACCCCGTGGCCGCTGGGAGTGTGGGAGTGTGGGAGTGTGGGAGTGTGGG 1247  
QY 1322 GCAGTCCGGGACAGCTCCAGTGTGAAGCTTCAAGTGTGTTCTTGAGACTGTGTATCC 1381  
Db 1248 GAAGCTTCCGGCTGAGCTCCAGTGTGAAGCTTCAAGTGTGTTCTTGAGACTGTGTATCC 1307  
QY 1382 AGAATCGATGTGCTTGAGGACAGGCTGGCTTCTTGGGATGTCTCCAGAACAAAGACT 1441  
Db 1308 AGAATCGATGTGCTTGAGGACAGGCTGGCTTCTTGGGATGTCTCCAGAACAGGCT 1367  
QY 1442 AACGACTACTGCTTTGATATTAACCTCCCGATGAAACCAAGATTTGGGACACAGGT 1501  
Db 1368 AAGAGGCTACTGCTTGAATATTAACCTCCCGATGAAACCAAGATTTGGGACACAGGT 1427  
QY 1502 CATTCGTACTCTGCTGATGAGTGGGAGGCTGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAG 1561  
Db 1428 CTTCTGTACTCTGCTGATGAGTGGGAGGCTGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAG 1487  
QY 1562 AGAATACGCTTATTAACACCCAGGCTGAGGCTGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAG 1621  
Db 1488 AGAATACGCTTATTAACACCCAGGCTGAGGCTGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAG 1547  
QY 1622 TACCCCTTATCATGATCTCTGCGAAGAAACTGCCCCAGAGAAATCAGAAAGTTCATCTTTGCA 1681

Db 1548 TACCCCTTATCATGATCTCTGCGAGAGACCGTCCCAGAGAACCCAGGAGTTCATCTTACA 1607  
QY 1682 GGAGGATGGATCTTTATTTTCCAGAACAGTCCAAAGAAATGTGTCCAGGCTCGGAGGAAGGA 1741  
Db 1608 GGAGGATGGACCTTTAGTTTCAAGACACAGCAGGAAATGTGTGGAGGCCACAGAGAGGT 1667  
QY 1742 GTCGAGTGACAGTTCGTTCCACTCTTTCAGAGACTGCACCAACTCCGATCATCAGAAATG 1801  
Db 1668 GTTAGACAACCGCTTTGACACACTTACGGGACTGTACCAACTCAGATAACCAAGAGGTG 1727  
QY 1802 GTTCTTCAAGAGGAGTGTATGAAGCTCTGTGTATCAAGGAGCCCATCGAAGGAGACT 1861  
Db 1728 GTTCTTCAAGGAGGAGT 1779  
QY 1862 GTGAGGAGGAGT 1921  
Db 1780 TCCAGGCTGTGACGT 1805  
QY 1922 ACTAGGCTGATCTGCTTTTGAAGAGGCAATCATTTTGCATTTGTGAAGTGTGTGTGTGT 1981  
Db 1806 -----TGAGATGGACAGCGTACACGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1850  
QY 1982 TTAGTAAATCTGAATAAGCTTTTGTACTTTATTTTGAAGAACTTTTAAATGTTCCTTCT 2041  
Db 1851 GTCAAGAGTGTGAATAAGCTTTTGTACTTTTGTACAGATTTTGAAGAACTTTTAAATGTTCCTTCT 1910  
QY 2042 ACCATTTTCAAGAGGTAATCGTAGATGTCTAACCTTGTATTTTGAAGAACTTTTAAATGTTCCTTCT 2101  
Db 1911 ATCTGCTTTCTGAGGCAATCTTAGATGTG-CCTTTGTGTGTGTGTGTGTGTGTGTGTGTGT 1962  
QY 2102 TTATAATTTTCTATCAARWAWATTTTACAGTCTGTCGCTTTTACTCTCATAGCA 2161  
Db 1963 -----ATATTTTCTTCAAGATGTGCTTTTACAACTGGTGTGCTTTTATTTCTCATAGCA 2017  
QY 2162 AAAAAAGATAAGATTTTATTTTGGTATTTTACAGAAATTTCCAGGTACAGAAATCTGCA 2221  
Db 2018 AAAAAAGATAAGATTTTATTTTGTATACATAAAGCTTCC----- 2057  
QY 2222 TGGGTGGAATCAGGTTTCAAGCAACGTTTCTGATTTTCACTGATTAATACCTCAGTCGG 2281  
Db 2058 -----TGGGTGTTTGTAGACATTTTACAACTGTTAATACCTCMAATGTTG 2106  
QY 2282 GGTAAAGTTTTCCTAGTATAGAGAGCTGCTACTAGGAACTTGTATTTGATTT-----A 2336  
Db 2107 AGTTAGATTTCTCTCAGGCGGATGAGTGGACACTAGAAATCTGTTGATTTCTTCAGGT 2166  
QY 2337 TTAGGCTTATGAGATCTTCTAGATGATTTTAAAAAGAAATGCTTTT 2383  
Db 2167 GTGAGTGGGTGAGATGATCTAGATTCATACATAAAGAAATGCTTTT 2213

RESULT 2  
AK042133  
LOCUS  
DEFINITION  
Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched library, clone:A630062B03 product:hypothetical Glycosyl transferase, family 2 containing protein, full insert sequence.  
AK042133  
VERSION  
AK042133.1 GI:26334956  
KEYWORDS  
HTC; CAP trapper.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus  
REFERENCE  
1  
Carninci,P. and Hayashizaki,Y.  
TITLE  
High-efficiency full-length cDNA cloning  
JOURNAL  
Meth. Enzymol. 303, 19-44 (1999)  
MEDLINE  
99279253  
PUBMED  
10349636  
REFERENCE  
2  
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,





## AUTHORS

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Konno, H., Akiyama, J., Niehi, K., Kiteunai, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,  
Yuneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
20530913  
11076861  
4

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,  
Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,  
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I.,  
Saito, T., Okazaki, Y., Gajobori, T., Bono, H., Kasukawa, T., Saito, R.,  
Kadota, K., Matsuda, H., Asburner, M., Batalov, S., Casavant, T.,  
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,  
Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Tonita, M.,  
Quackenbush, J., Schriml, L.M., Staab, F., Suzuki, R., Tomita, M.,  
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,  
Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,  
Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C.,  
Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,  
Hofmann, M., Hume, D.A., Kaniya, M., Lee, N.H., Lyons, P.,  
Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P.,  
Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,  
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H.,  
Toyo-oka, K., Wang, K.H., Weitz, C., Whittaker, C., Wilming, L.,  
Wynshaw-Boris, A., Yoshida, K., Haegawa, Y., Kawaji, H., Kohtsuki, S.  
and Hayashizaki, Y.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409 (6821), 685-690 (2001)  
21085660  
11217851  
5

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

The FANTOM Consortium and the RIKEN Genome Exploration Research  
Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
6 (bases 1 to 2034)

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,  
Fukuda, S., Furuno, M., Hasegawa, T., Hara, A., Hashizume, W.,  
Hayashida, K., Hayata, K., Hiramoto, K., Hirose, T., Hirozane, T.,  
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,  
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,  
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,  
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,  
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,  
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
Muramatsu, M. and Hayashizaki, Y.  
Direct Submission

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,  
URL: <http://genome.gsc.riken.go.jp/>, Tel. 81-45-503-9222,  
Fax: 81-45-503-9216)

## COMMENT

cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site for further details.  
URL: <http://genome.gsc.riken.go.jp/>  
URL: <http://fantom.gsc.riken.go.jp/>  
Location/Qualifiers

## FEATURES

## source

1. .2034

```

/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="PANTOM DB:C130094P07"
/db_xref="taxon:10090"
/clone="C130094P07"
/tissue_type="head"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="16 days embryo"
misc_feature
66..1797
/notes="hypoetical Glycosyl transferase, family 2
containing protein (InterPro|IPR001173, evidence:
InterPro)
putative"
BASE COUNT 439 a 528 c 635 g 432 t
ORIGIN
Query Match 44.1%; Score 1257; DB 11; Length 2034;
Best Local Similarity 82.6%; Pred. No. 8.4e-128;
Matches 1480; Conservative 0; Mismatches 295; Indels 16; Gaps 3;
QY 54 AGATCGCTGGCTGCAGTTGGCGGGCGCATGTGGGGGGGACGCGGGCGGGCGCGCTGCCG 113
DB 39 AGCCCGCGGGGTGCAGCTAGCGGGTGCATGTGGGGGGCGTGGGTGAGGAGCGCGTGGCCG 98
QY 114 CGGGAACCTGGCGGCGCGCGGGGAGGCGCTGTGGTGTCTCTCGCTGCTACTTGGCGTTGGCC 173
DB 99 CGGGGGCTGGCGGCGCGCGGGGAGGCGTTG-----CTGGCACTGCTGGCGCTGGCA 149
QY 174 GGGCTGGGCTCGTGTGTGGCGGCGCAGCGTGGGGCGGGCGGGGGTGGCGAGCGCGGA 233
DB 150 GGGCTGGGCGCGCTGCTGGCGGGC-----TCGAAGCAGGTCGGGGACGGTGGACCCCGGA 203
QY 234 CCCCCGCGACCCCGCGCGCGCGCGCGCGCGCGCGTGTATGCGCGGGCGCGCGCTGCCG 293
DB 204 CCCCCGCGGACTCTCTCTCCCGCGCGCTACAGCCAGTGTCTGCCAGCGCGCGCTGCC 263
QY 294 GCGAACCGCTGGCGCGCGCGGGCGGCTGGCTGCGAGCTGCGAGGCGGAGGAGCTG 353
DB 264 GCGATGTCTTGGAGCGCATGGGAGCGGTGGTGGCTGCGAGCGGAGGAGCTG 323
QY 354 CGGCTGCAGGAGGAGCGTGGCGGTGCACAGATTAACTTACCTCCTCAGCGAGCGCATC 413
DB 324 CGGCTGCAGGAGGAGCGGTGAAGCAGCACCAGATCAATCTATCTTGAGCGACCGTATC 383
QY 414 TCAGTGCACCGCGCTGCCCGCGCGCTGAAACCGCTGAAACCGCTGCAAGAGAGAAATATGAT 473
DB 384 TCGCTGCACCGCGCTGCCCGCGCGCTGAAACCGCTGAAACCGCTGCAAGAGAGAGTAAATATGAT 443
QY 474 TATGATAATTTGCCAGGACATCTGTTAT-CATAGCAATTTTATAATGAAGCTGCTCAAC 532
DB 444 TATGATAATTTGCCAGGACATCTGTTATGTCATAGCAATTTTATAATGAAGCTGCTCAAC 503
QY 533 TCTCTTTTCGACAGTGTTCAGTGTCTTTGAGACATCCCGGATATCTCTGTAGAGAGT 592
DB 504 ACTCTTTTCGACAGTGTTCAGTGTCTTTGAGACATCCCGGATATCTCTGTAGAGAGT 563
QY 593 GATCTTGTAGTACGTACAGTGTAGAGCAGCTTGAAGGCGCTTGGCCATGAGCT 652
DB 564 CATCTGTGTAGTACGTACAGTGTAGAGCAGCTTGAAGGAGCGCTTGGCCATGAGCT 623
QY 653 TTGGGAGTCTCCCAAGGTGGCGCTGTATGTCGCGCGCAACAAGAGAGAGGCGCTTGGTGGCAGC 712
DB 624 GTACAGCTCTCCCAAGGTGGCGCTGTATGTCGTCAGAGAGAGAGGCGCTTGGTGGCAGC 683
QY 713 CCGGCTGTGGGGCGCTTGGCGGAGGCGAGTGTCTGACCTTCTTGGACTGTCTACTG 772
DB 684 CCGGCTGTGGGGCGCTTGGCGGAGGCGAGTGTCTGACCTTCTTGGACTGTCTACTG 743
QY 773 TGAATGTATGAGGAGTGGCTGGAGCGCTGCTGTCAGAGGATCCATGAGAGGAGTGGC 832
DB 744 TGAATGTATGAGGAGTGGCTGGAGCGCTGCTGTCAGAGGATCCATGAGAGGAGTGGC 803

```





Db 246 GGTAGTAGCTACAGCGACAGAGACACCTAAAGAACGCTTGGCCAAAGAGCTGTCA 305  
Qy 659 ACTGCCCAAGGTGGCTGTATCCGCGCAACAAGAGAGAGGGCTGTGGAGCCCGGCT 718  
Db 306 GCTCCCAAGGTGGCTGTATCCGCGCAACAAGAGAGAGGGCTGTGGAGCCCGGCT 365  
Qy 719 GCTGGGGGGCTGTGGCGGAGGGCGATGTTCTGACCTTCTTGACCTGTCTGAGTG 778  
Db 366 ACTGGGAGCTCTCGCGCCAGGGCGAAGTGTCTGACCTTCTTAGACTGTCTGATG 425  
Qy 779 CCACGAAGGGTGGCTGGAGCGCTGTCTGACAGAGATCCATGAAGAGAGTGGCAGTGT 838  
Db 426 TCATGAGGGTGGCTGGAGCGCTGTCTGACAGAGATCCACGAGAGGAGTCCGAGTGT 485  
Qy 839 GTGCCCGGTGATGTGATCGATCGAACAACCTTCGAATACCTGGGGAACCTCCGGGA 898  
Db 486 GTGCCCTGTATCGATGTGATGTAAGTGAACACCTTTGAGTACCTGGGCAACTCGGGA 545  
Qy 899 GCGCCAGATGGCGGCTTTCAGCTGGAGCTGGTGTTCACGTGGGCACACAGTTCCTGAG 958  
Db 546 GCGCCAGATGGCGGCTTTCAGCTGGAGCTGGTGTTCACGTGGGCACACAGTTCCTGAG 605  
Qy 959 GGAGAGATACGATGCAATCCCGCTCGATGTCTCAGGTCTCCAAATGGCTGGTG 1018  
Db 606 GGAGCGAGTCGATCGGCTGGCGATGATGTTATCAGGTCTCCAACTATGGCTGGGG 665  
Qy 1019 GCTGTTGCTGTGATGAAGAAATATTTGAAATATCTGGGTCTTATGATACAGAAATGGA 1078  
Db 666 CTTGTTGCTGTGATGAAGAAATATTTGAAATATCTGGGTCTTACGATACAGAAATGGA 725  
Qy 1079 AGTTGGGGAGGAAACCTCGAATTTCTCTTTAGGATCTGGCAGTGGTGGGTTCT 1138  
Db 726 AGTCTGGGGAGGAAACCTTGAGTTCTCTTTAGGATCTGGCAGTGGTGGGCACTCT 785  
Qy 1139 GGAACACACACCATGTTCCCATGTGGCCATGTTTCCCAAGCAAGCTCCCTACTCCG 1198  
Db 786 GGAACGACCCGCTGCTCCCATGTGGCCAGCTCTCCCTAAGCAAGCTCCCTATTCG 845  
Qy 1199 CAACAAGCTCTGGCCAAAGTGTCTGACGTGAAGTATGATGATGAATTAAGA 1258  
Db 846 TAGCAAGCCCTGGCCAAAGTGTCCGAGCTGCAGAAAGTGTGATGAATTAAGA 905  
Qy 1259 GCTCTACTACATCGCAACCCCGCTGGCGCTTCCGAACCTTTTGGGATGTGACAGAG 1318  
Db 906 ACTCTACTACACCCGCAATCCCGAGCCCGCTGGAACTTTGGGAGCTGACAGAG 965  
Qy 1319 GAAGCAGCTCCGGCAAGCTCCAGTGTAAAGACTTCAAGTGGTCTTTGGAGACTGTA 1378  
Db 966 GAAGAGCTCTGGGCTAAGCTCCAGTGTAAAGACTTCAAGTGGTCTTTAGATACAGTGA 1025  
Qy 1379 TCCAGAACTGATGTGCTGAGGACAGCCCTGGCTTCTTCGGGATGCTCCAGAAACAAG 1438  
Db 1026 CCCAGAAATGCAAGTGCACAGGACAGCCCTGGCTTCTTTGGGATGCTCCAGAAACAAG 1085  
Qy 1439 ACTAACAGACTCTGCTTGTGACTAATCCCTCCGATGAACACAGATGTTGGGACACCA 1498  
Db 1086 CCTAAGAGGACTGCTGCTGTGACTAATCCCTCCGATGAACACAGATGTTGGGACACCA 1145  
Qy 1499 GGTCACTCTGACTCTGCTGATGGATGGGCGAGAAATCAGTGTTCGAGTACAGTCCCA 1558  
Db 1146 GGTCTCTGTAACCTCTGCTGATGGATGGGTCAGAACCCAGTGTTCGAGTATACGACCG 1205  
Qy 1559 GAAAGAAATACGCTATAAACCCACCGCTGAGGCTGCAATGCTGTGGAGCAGGAAT 1618  
Db 1206 GAAAGAAATACGCTATAAACCCACCGCTGAGGCTGCAATGCTGTGGAGCAGGAAT 1265  
Qy 1619 GGATACCTTATCATGCTCTCTGCGAAGAACTGCCCCAGAGATCAGAAAGTTCATCTT 1678  
Db 1266 GGATACCTTATCATGCTCTCTGCGAAGAACTGCCCCAGAGACCGTCCAGAGAACCGAGGATTCATCT 1325  
Qy 1679 GCAGGAGGATGGATCTTTTATTTTACGAACAGTCCAAAGAAATGTGTCCAGGCTGGAGGAA 1738  
Db 1326 ACAGGAGGATGGATCTTTTATTTTACGAACAGTCCAAAGAAATGTGTGGAGGCCACAGAGAA 1385

Qy 1739 GGAGTCTGAGTGCAGTTCGTTCCACTCTTACGAGACTGCACCAACTCGGATCATCAGAA 1798  
Db 1386 GGTGTTAGACAACGGCTTTGACCACTACTTACGGAGCTGTACCACTCAGATAACCCAGAG 1445  
Qy 1799 ATGGTTCCTTCAAGAGCGCATGTTATGAAGCCTCGTGTATCAAGGAGCCATCGAAGAGGAG 1858  
Db 1446 GTGGTTCCTTCAAGGAGCGCATGTCGTAGTGCATCGCT-----TGCCAAGGAAAGGAG 1497  
Qy 1859 ACTGTGGAGCCAGGACTCTGCCCAACAAGACTTAGCTAAGCAGTGCACGACCCACCA 1918  
Db 1498 CCTCCACGGCTGTGGACTGTGTGCCAA----- 1526  
Qy 1919 AAAAATAGCTGCATGCTTTTGAAGAGGCAATCATTTTGCCATTTGTGAAAGTGTGTG 1978  
Db 1527 -----TGAGATGGACAGCGTACACGGTGTCTTCGTGAA---TGTCGTG 1568  
Qy 1979 GATTTAGTAAAAATGTAATAAGCTTTGTACTTATTTTGAAGAACTTTTAAATGTTCCTCAA 2038  
Db 1569 GGTGTCAAGAAGTGTGAATTAAGCTTTGTACAGATTTTGAAGAACTTTTAAATGTCTCCT 1628  
Qy 2039 AATACCTATTTTCAAGGGTAATCTGAAGATGTTAACCTTGGTATTTAGAAATTA 2098  
Db 1629 GGCATCTCTGCTTCTGAGGGCAATCTTAAGATGTT-GCCTTTGGTATTTGGAA----- 1680  
Qy 2099 ACCTTATAATATTTTCTATCAARAWAWATTTTACAGTCTGTCCTTTTACTCTCATTA 2158  
Db 1681 -----AGGATATTTTCTTACAGATGTCATTTTACATGTCCTTTTACTCTCATTA 1735  
Qy 2159 GCAAAAAAGATAAGATTTTATTTTGGTATTTTACAAGAA 2197  
Db 1736 GCAAAAAAGTAAATACTTTTATTTTGCATACAAAAAAA 1774

RESULT 5  
BC024475

LOCUS Mus musculus, clone IMAGE:4976620, mRNA. linear HTC 04-MAR-2003  
DEFINITION BC024475  
ACCESSION BC024475.1 GI:22137472  
VERSION BC024475.1  
KEYWORDS HTC.

SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 2064)  
AUTHORS Strausberg, R.  
TITLE Direct Submission  
JOURNAL Submitted (01-MAR-2002) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA

REMARK NTH-MGC Project URL: <http://mgc.nci.nih.gov>  
COMMENT Contact: MGC help desk

Email: [cgapbs-1@mail.nih.gov](mailto:cgapbs-1@mail.nih.gov)  
Tissue Procurement: Jeffrey Green M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Sequencing Group at the Stanford Human Genome  
Center, Stanford University School of Medicine, Stanford, CA 94305  
Web site: <http://www-shgc.stanford.edu>  
Contact: (Dickson, Mark) [mcdpaxil.stanford.edu](mailto:mcdpaxil.stanford.edu)  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,  
R. M.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAC Plate: 59 Row: j Column: 15  
This clone was selected for full length sequencing because it  
passed the following selection criteria: Hexamer frequency ORF  
analysis  
This clone has the following problem: retained intron.  
Location/Qualifiers

```

source      1. .2064
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4976620"
/tissue_type="Mammary tumor, C3(1)-Tag model. Infiltrating
ductal carcinoma, 5 month old virgin mouse."
/clone_lib="NCI CGAP Mam6"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"

BASE COUNT      475 a   505 c   625 g   459 t
ORIGIN
Query Match      34.8%; Score 992; DB 11; Length 2064;
Best Local Similarity 69.8%; Pred. No. 4.8e-99;
Matches 1626; Conservative 4; Mismatches 409; Indels 291; Gaps 10;

QY      59  GCTGGCTGCAGTTGGCGGGCGCATGTGGGGCGGCACGGCGGGCGCGCTGCCCGCGGGA 118
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB      1  GCGGGGTGCAGCTAGACGGGTGCATGTGGGGCGGTGCGGTGAGGAGGCGCTGCCCGCGGGG 60
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY     119  ACTGCGGCGCGCGGAGGCGCTGTTGTGCTCTCTGGCGCTACTGTCGGTGTGGCGGGCT 178
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB      61  GCTTGGCGCGCGCGGAGGCGTTG-----CTGGCACTGCTGGCCCTTGGCAGGGCT 111
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY     179  GGGCTCGGTGCTTGGCGGCGCAGCGTGGGGCGGGCGTCCCGAGCGGGACACCCC 238
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB     112  GGGCGCGCTGCTGCGGCG-----TCGAAGCAGGTGCGGGACGGTGACCCGGGACCCCC 165
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY     239  GCGGACCCCGCGCCCGGGCGGCGAGCGGTGATGCCGGCGCGCGCGTGC CGGCGGAA 298
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB     166  GCGGACTCTCTCTCCCGGGGTCACGAGCAGTGTCTGCCAGGCGCGCTGCCCGCGCGA 225
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY     299  CGCGCTGGGCGCGGGCGAGCGGTGCGGCTGCAGCTGCAGGGCGAGAGCTGCGGCT 358
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB     226  TGTCTGGGAGCGCATGGGAGGCGGTGCGGCTGCAGCTGCAGGGCGAGAGCTGCGGCT 285
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY     359  GCAGGAGGAGCGTGGCGGTGCACCAAGATTAACTCTACTCAGCGACCGCATCTCACT 418
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB     286  GCAGGAGGAGCGTGNAGCAGACCAGATCAACATCTATCTGAGCGACCGTATCTCGCT 345
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY     419  GCACCGCGCTGCCCGAGCGCTGGAAACCCGCTGTGCAAAAGAGAAATATGATTATGA 478
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB     346  GCACCGCGGACTGCCCGAGCGCTGGAAACCCCT----- 378
      |||  |||  |||  |||  |||  |||  |||  |||  |||

QY     479  TAATTTGCCCAGGACATCTGTTATCATAGCATTTTATAATGAAGCTGGTCAACTCTCT 538
      |||  |||  |||  |||  |||  |||  |||  |||  |||
DB     379  ----- 378
      |||  |||  |||  |||  |||  |||  |||  |||

QY     539  TCGGACAGTTTACAGTGTCTTGAGACATCCCGGATATCTGTAGAGAAGTGATCTCT 598
      |||  |||  |||  |||  |||  |||  |||  |||
DB     379  ----- 378
      |||  |||  |||  |||  |||  |||  |||

QY     599  TGTAGATGACTACAGTGATAGAGAGCACCTGAAGAGCGCTTGCGCCAATGAGCTTTCGGG 658
      |||  |||  |||  |||  |||  |||  |||  |||
DB     379  -----AGCACCTAAAGGAACGCTTGGCCAACAGAGCTGTCA 415
      |||  |||  |||  |||  |||  |||  |||

QY     659  ACTGCCCAAGTGGCGCTGATTCGCGGCCAACAGAGAGAGGCGCTGTGCGAGCCCGGCT 718
      |||  |||  |||  |||  |||  |||  |||  |||
DB     416  GCTCCCCAAGTGGCGCTGATTCGCTGTAGCAGGAGAGGCGCTAGTGGGAGCCCGGCT 475
      |||  |||  |||  |||  |||  |||  |||

QY     719  GCTGGGGCGCTGCGGCGAGGGGCGATGTTCTGACCCTTCGAGCTGTCACCTGTGAGTG 778
      |||  |||  |||  |||  |||  |||  |||
DB     476  ACTGGGAGCCTCTGCGGCCAAGGGCGAAGTGTGACCTTCTTAGCTGTCTACTGTGAATG 535
      |||  |||  |||  |||  |||  |||

QY     779  CCACGAAGGTTGGTGGAGCGGCTGCTGCGAGAGGATCCATGAAGAGAGTCCGCGAGTGGT 838
      |||  |||  |||  |||  |||  |||
DB     536  TCATGAGGGTGGCTGGAGCCTCTGCTGCGAGAGATCCAAGAGAGAGTCCGCGAGTCTGT 595
      |||  |||  |||  |||  |||

QY     839  GTCCCGGTGATTGATGTGATCGACTGGAAACA CTTTCGAATATCCTGGGGAATCCCGGGA 898
      |||  |||  |||  |||  |||
DB     596  GTCCCTGTTATCGATGTGATTGACTGGAACA CACTTTGATGACTCGGCAACTCGGAGA 655
      |||  |||  |||  |||  |||

```

```

Qy 1979 GATTAGTAAAGTGTGAATAAGCTTTGTACTTATTTTGTAGAACTTTTAAATGTTCCAA 2038
Db 1679 GGTGTCAACAGAGTGTGAATAAGCTTTGTACAGATTTTGAAGAACTTTTAAATGTTCTCT 1738
Qy 2039 AATACCCCTATTTTCAAGGGTAATCGTAAGATGTTAACCCCTGTGTATTTAGAAATTTAA 2098
Db 1739 GGCATCTCTGCTTTCTGAGGGAATCTTAAAGATGTT-GCCTTTGTGTATTTGGAA----- 1790
Qy 2099 ACCTTATAATATTTTCTATCAARAWAWATTTTACAGTGTCTGCTTTACTCTCTCATTA 2158
Db 1791 -----AGATATTTTCTTATCAAGATGTCATTTTACAATGTCCTTTTATTTCTCATTA 1845
Qy 2159 GCAAAAAGATAAAGATTTTATTTTGTGTATTTTACAAGATTTTCCACAGGTACGAAGATATCT 2218
Db 1846 GCNAAGAGGTAAATACTTTTATTTTGTATACATACACAAAGCTTCTCTGGTGTT----- 1897
Qy 2219 GCATGGGTGGAATCAGTTTCAACAGCTGTACAGTGTCTGATGATGTTTAAAGAAATGCTTTT 2383
Db 1898 -----GTAGCACATTTTACACAAACTGTTAAATACCTTCAAAATG 1934
Qy 2279 CGGGGTTAAAGTTTCCAGTATAGAGAGCTGTCACTAGGAACATTTGTTGATTT--- 2335
Db 1935 TTGAGTTAGATTTCTCTCAGGGGATGAGTGGACACTAGAAATACTGTTGATTTCTTCA 1994
Qy 2336 --ATTTCAGGTCAATTCAGATCTTCTAGATGTTATTTTAAAGAAATGCTTTT 2383
Db 1995 GGTGTGAGTGGGTGAGATGCTCTAGATTCATACATACAAAGAAATGCTTTT 2044

RESULT 6
LOCUS AK043977
DEFINITION Mus musculus 10 days neonate cortex cDNA, RIKEN full-length enriched library, clone:A830062E06 product:hypothetical Glycosyl transferase, family 2 containing protein, full insert sequence.
ACCESSION AK043977
VERSION AK043977.1 GI:26090099
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
MEDLINE 10349636
REFERENCE
2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
MEDLINE 11042159
REFERENCE
3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
MEDLINE 11076861
REFERENCE
4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I.,

```

```

Saio, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pegole, G., Quackenbush, J., Schriml, L. M., Staubli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, F., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S. and Hayashizaki, Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)
21085660
MEDLINE 11217851
REFERENCE
5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 1752)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, K., Saiton, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
COMMENT
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/
FEATURES
Location/Qualifiers
1..1752
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:A830062E06"
/db_xref="taxon:10090"
/clone="A830062E06"
/tissue_type="cortex"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="10 days neonate"
757..1612
/notes="hypothetical Glycosyl transferase, family 2 containing protein (InterPro|IPR001173, evidence: InterPro) putative"
misc_feature
371 a 468 c 566 g 347 t
BASE COUNT
ORIGIN

```

Query Match	33.1%;	Score 943.8;	DB 11;	Length 1752;
Best Local Similarity	74.4%;	Pred. No. 8.8e-94;		
Matches 1333;	Conservative 0;	Mismatches 27;	Indels 186;	Gaps 4;
Qy	54	AGATCGCTGGCTGCAGTTGGCGGGCGCATGTGTGGTTCCTTGGGCGTACTTGGCGTTGGCC	113	
Db	24	AGGCCGGGGTGCAGCTAGCGGGTGCAATGTGGGGCGTGCGGTGAGAGAGCGCTGCCCG	83	
Qy	114	CGGAAACTGCGCGCGCGCGCGGAGGCGCTGTGTGGTTCCTTGGGCGTACTTGGCGTTGGCC	173	
Db	84	CGGGGGCTGCGCGCGCGCGGAGGCGTTG-----CTGGACTCTCTGGCCTTGCCA	134	
Qy	174	GGGCTGGGCTCGGTCTGTCGCGGCGACGCTGGGGCGGGGCTGCCGAGCCGGGA	233	
Db	135	GGGCTGGGCGCGCTGCTCGGGC-----TCGAAGCAGTTCGGGACGTTGGAACCGGGA	188	
Qy	234	CCCCCGCGCACCCCGCGCCCCCGGGGGCGAGCCGGTCAATGCCGCGGCCCGCGTGGCG	293	
Db	189	CCCCCGCGGACTCTCTCTCCCCCGGCGTCAAGAGCAGTGTCTGCCAGGCGCGCTGCC	248	
Qy	294	GCGAACCGCTGGCGCGCGCGGCGAGSCGTGGCGGTGCAGCTGCAGGGCGAGGAGCTG	353	
Db	249	GCCGATGCTCTGGAGAGCGCATGGGAGGCGGTGCGCTGCAGCTGCAGGGCGAGGAGCTG	308	
Qy	354	CGGCTGCAGGAGGAGCGTGCCTGTGCACAGATTAACATCTACTCAGCGACCGCATC	413	
Db	309	CGGCTGCAGGAGGAGCGTGAAGCAGCACAGATCAACATCTATCTGAGCGACCGTATC	368	
Qy	414	TCACTGCACCGCCGCTGCCCGAGGCTGGAAACCGCTGTGCCAAGAGAGAAATATGAT	473	
Db	369	TCGCTGCACCGCCGACTGCCCGAGCGTGGAAACCTCT-----	406	
Qy	474	TATGATRAATTGCCAGGACATCTGTATCATGCAATTTTATATGAAGCCTGGTCAACT	533	
Db	407	-----	406	
Qy	534	.CTCCTTCGGACAGTTTACAGTGTCTTTGAGACATCCCGGATATCTCTAGTAGAAGTG	593	
Db	407	-----	406	
Qy	594	ATCCTTTAGATGACTACGTGATAGAGAGCACCTGAAGAGCGCTTGGCCAAATGAGCTT	653	
Db	407	-----AGCACCTAAAGGAACGCTGGGCCAACGAGCTG	438	
Qy	654	TCGGGACTGCCAAGCTGCCCTGATCCGCGCCAAACAGAGAGGCGCTGGTGGAGCC	713	
Db	439	TCAAGCTCCCCAAGTGCCTGATCCGTGCTAGCAGGAGAGGCGCTAGTGCAGGCC	498	
Qy	714	CGGCTGTGGGGGCTGTGCGCGCAGGGGCGATGTTCTGACCTTCTGGACTGTCACTGT	773	
Db	499	CGGCTACTGGAGCCTCTGCGCCAGGGCGGAGTGTGACCTTCTTAGACTGTCACTGT	558	
Qy	774	GAGTGCCAGAAAGGTGGCTGGAGCCGCTGTCAGAGGATCCATGAAGAGGAGTCCGCA	833	
Db	559	GAATGTCTAGGGGTGGCTGGAGCCTCTGCTGCAGAGGATCCACAGAGGAGTCCGCA	618	
Qy	834	GTGGTGTCCCGGTGATGTGATCGACTGGAAACACTTTCGAATACCTTGGGGAACTCC	893	
Db	619	GTCTGTGCCCTGTATTTCGATGTGATTGACTTGGAAACACTTTGAGTAGTACCTGGGCAACTCG	678	
Qy	894	GGGAGAGCCAGATCGGCGTTTCGACTGGAGGCTGGTTCAGCTGGGCACACAGTTTCT	953	
Db	679	GGAGAGCCGAGATTGGCGGCTTTGACTGTGGCGGTGGTATTTCAGTGGCAGCTGGTCCC	738	
Qy	954	GAGAGGGAGAGGATACCGGATGCAATCCCCCGTCCGATGTCTACAGTCTTCCAAACATGGCT	1013	
Db	739	CAGCGGAGCGGCACTCGATGCGGTTCGCGGATCGATGTTATCAGGTCTTCCAACTATGGCT	798	
Qy	1014	GGTGGGCTGTTGTGTGAGTAAAGAAATATTTTGAATATCTGGGGTCTTATGATACAGGA	1073	
Db	799	GGGGGCTGTTGTGTGAGTAAAGATATTTGATTTATCTGGGGTCTTACGATACAGGA	858	

QY	1074	ATGGAAGTTGGGGAGGAGAAAACCTTCGAAATTTTCTTTAGAGTCTGGCAGTGTGGTGGG	1133
Db	859	ATGGAAGTCTGGGGAGGAGAAAACCTTTGAGTTCTCTCTTTAGGATCTGGCAGTGTGGTGGC	918
QY	1134	GTTCCTGGAACACACCCATGTTCCCATGTTTGGCCATG-TTTTCCCCCAAGCAAGCTCCCTA	1192
Db	919	ACTCTGGAACAGCACCCGTGCTCCCATGTGGGCCACGCTCTTCCCTAAGCAAGCTCCCTA	978
QY	1193	CTCCCGCAACAAGGCTCTGGCCCAACAGTGTTCGTGACAGTGAAGTATGGATGGATGAATT	1252
Db	979	TTCCCGTGGCAAGGCCCTGGCCCAACAGTGTCCGAGCTGCAGAGTGTGGATGGATGAATT	1038
QY	1253	TAAAGAGCTCTACTACATCGCAACCCCGTGGCCGCTTGGAAACCTTTTGGGGATGTGAC	1312
Db	1039	TAAAGAACTCTACTACCACCGCAATCCCAAGGCCCGCTGGAACCCCTTTGGGGACGTGAC	1098
QY	1313	AGAGAGAAGCAGCTCCGGGACAGCTCCAGTCTGAAGACTTCAAGTGGTCTTTGGAGAC	1372
Db	1099	AGAGAGAAGAGCTTCGGGCTAAGCTTCCAGTGTGAAGACTTCAAGTGGTCTTTAGATAC	1158
QY	1373	TGTGTATCCAGAACTGCATGTGCTGAGGACAGGCTTGGCTTCTTCGGGATGCTCCAGAA	1432
Db	1159	AGTGATCCACAGATTGCACTGTCGAGGACAGGCTGGCTTCTTTGGGATGTCCAGAA	1218
QY	1433	CAAGAAGCTAACAGACTACTGCTTTGACTATACCCCTCCGATGAAACACAGATTTGTGG	1492
Db	1219	CAGAGGCTTAAGAGGGTACTGCTTGACTACAATCCTCCCAATGAAACCATGTTGAAG	1278
QY	1493	ACACCAAGTCTATCTGTAAGTCTGTATGGGATGGCCAGAAATCAGTTTTTCAGATACAC	1552
Db	1279	CCACCAAGTCTTCTGTGTAAGTCTGTGCAATGGGATGGGTGCAACCAAGTTTTTCAGATAC	1338
QY	1553	GTCCCAAGAAAGAAATACGCTATAACCCACACGAGCTGAGGGGTGCAATGCTGTGAAGC	1612
Db	1339	GACCCGAAAGAAATACGCTATAACCCGCGAGCCGGAAGCTGTATCACCTGTGAGGA	1398
QY	1613	AGGAATGGATACCCCTTATCATGATCTCTGCGAAGAACTGCCCCAGAGAAATCAGAAGTT	1672
Db	1399	CGGAAGGATACCCCTTGTGATGATCTCTGCCGAGAGACCGTCCCAGAGAAACAGGAGTT	1458
QY	1673	CATCTTGCAGGAGATGGATCTTTATTTACGAAACAGTCCAAGAAATGTGTCCAGGCTGC	1732
Db	1459	CATCCTACAGGAGGATGGCACTTTAGTTTCAAGACAGCAGGAGAAATGTGTGAGGCCAC	1518
QY	1733	GAGGAAGGAGTCTGAGTGACAGTTTTCGTTCCACTCTTTACGAGACTGCACCAACTCGGATCA	1792
Db	1519	AGAGAAGGTGTAGACACAGGCTTTGACCATACTTACGGGACTGTACCAACTCAGATAA	1578
QY	1793	TCAGAAATGTTCTTCAAAGAGCGCATGTTATGAAGCCTCGTGTATCAAGG	1843
Db	1579	CCAGAGGTGTTCTTCAAAGGAGCGCATGTCGTAGTGCATCGCCTTGCCCAAG	1629
RESULT 7			
CD251293			
LOCUS		909 bp mRNA linear EST 22-MAY-2003	
DEFINITION		ACENCOURT 14211744 NIH MSC 179 Homo sapiens cDNA clone	
		IMAGE:30384460 5', mRNA sequence.	
ACCESSION		CD251293	
VERSION		CD251293.1 GI:31011759	
KEYWORDS		EST.	
SOURCE		Homo sapiens (human)	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
TITLE		1 (bases 1 to 909).	
JOURNAL		NIH-MGC http://mgc.nci.nih.gov/.	
COMMENT		National Institutes of Health, Mammalian Gene Collection (MGC)	
		Unpublished	
		Contact: Daniela S. Gerhard, Ph.D.	
		Office of Cancer Genomics	
		National Cancer Institute / NIH	
		Bldg. 31 Rm10A07 Bethesda, MD 20892	

Email: cgabs-r@mail.nih.gov  
 Tissue Procurement: Dr. Michael Brownstein  
 cDNA Library Preparation: Invitrogen Corp  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
 Plate: NDAM449 row: p column: 05  
 High quality sequence stop: 701.

FEATURES  
 source  
 1..909  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:30384460"  
 /tissue\_type="Pituitary"  
 /lab\_host="DH10B-Ton A (T1 and T5 phage resistances)"  
 /clone\_lib="NIH\_MGC\_179"  
 /notes="Organ: Brain; Vector: pCMV-SPORT6.1; Site 1: EcoRV  
 (destroyed); Site 2: NotI; Library is oligo-dT primed and  
 directionally cloned (EcoRV site is destroyed upon cloning  
 ). Average insert size 1.1 kb. Library was constructed by  
 (Invitrogen). Note: this is a NIH\_MGC Library."  
 BASE COUNT 273 a 178 c 198 g 260 t  
 ORIGIN  
 Query Match 30.5%; Score 868; DB 14; Length 909;  
 Best Local Similarity 98.3%; Pred. No. 2e-85;  
 Matches 871; Conservative 5; Mismatches 10; Indels 0; Gaps 0;  
 QY 1490 GGGACACAGGTCTTCTGACCTCTGATGGATGGCGGAGATCAGTTTCGAGTA 1549  
 DB 22 GGGATCCAGGTCTTCTGACCTCTGATGGATGGCGGAGATCAGTTTCGAGTA 81  
 QY 1550 CAGCTCCAGAGAAATAGCTATACACCCAGCTGAGGCTGCTGTTGGA 1609  
 DB 82 CAGCTCCAGAGAAATAGCTATACACCCAGCTGAGGCTGCTGTTGGA 141  
 QY 1610 AGCAGGAATGATACCTCTATCATGCTCTCTGGAAGAAATGCCCCAGAGATCAGAA 1669  
 DB 142 AGCAGGAATGATACCTCTATCATGCTCTCTGGAAGAAATGCCCCAGAGATCAGAA 201  
 QY 1670 GTTCATCTGAGGAGATGGATCTTTATTTTACGAAACAGTCCAGAAATGTCTCAGGC 1729  
 DB 202 GTTCATCTGAGGAGATGGATCTTTATTTTACGAAACAGTCCAGAAATGTCTCAGGC 261  
 QY 1730 TGGCAGGAGAGTGGAGTACAGTTTCTGCTTCTTACGAGCTGACCACTCGGA 1789  
 DB 262 TGGCAGGAGAGTGGAGTACAGTTTCTGCTTCTTACGAGCTGACCACTCGGA 321  
 QY 1790 TCATCAGAAATGGTTCTTCAAGAGCGCATGTTATGAAGCTCTGTATCAAGAGGCCA 1849  
 DB 322 TCATCAGAAATGGTTCTTCAAGAGCGCATGTTATGAAGCTCTGTATCAAGAGGCCA 381  
 QY 1850 TCGAAGGAGACTGTGGAGCCAGGACTGTGCCCAACAAAGACTTACCTAAGCAGTGCAG 1909  
 DB 382 TCGAAGGAGACTGTGGAGCCAGGACTGTGCCCAACAAAGACTTACCTAAGCAGTGCAG 441  
 QY 1910 AACCACCAAAACTAGGCTGCTTCTTCAAGAGGCAATCATTTTGGCAATTTGTGAAA 1969  
 DB 442 AACCACCAAAACTAGGCTGCTTCTTCAAGAGGCAATCATTTTGGCAATTTGTGAAA 501  
 QY 1970 GTTGTGTTGGATTTAGTAAAAATGTAAATAGCTTTGTACTTATTTTGGAACTTTTAA 2029  
 DB 502 GTTGTGTTGGATTTAGTAAAAATGTAAATAGCTTTGTACTTATTTTGGAACTTTTAA 561  
 QY 2030 ATGTTCCAAATACCTATTTTCAAAGGTAATCGTAAGATGTTAACCTTGGTATTAG 2089  
 DB 562 ATGTTCCAAATACCTATTTTCAAAGGTAATCGTAAGATGTTAACCTTGGTATTAG 621  
 QY 2090 AAAATTAACCTTATATATTTTCTATCAARAWAWATTTTACAGTCGTGCTTTTA 2149

Db 622 AAAATTAACCTTATATATTTTCTATCAAGATGTATATTTTACAGTCGTGCTTTTA 681  
 QY 2150 CTCTCATTAGCAAAAGATTAAGATTTTATTTTGTATTTTCAAGAAATTTCCAGGTACG 2209  
 Db 682 CTCTCATTAGCAAAAGATTAAGATTTTATTTTGTATTTTCAAGAAATTTCCAGGTACG 741  
 QY 2210 AAGATATCTGCATGGTGGAAATCAGGTTTCAAGCAACGTACTTTGCATTAACTGATAA 2269  
 Db 742 AAGATATCTGCATGGTGGAAATCAGGTTTCAAGCAACGTACTTTGCATTAACTGATAA 801  
 QY 2270 CTTCAAGCTCGGGTTAAAGTTTCCAGTATAGAGAGACTGTCTACTAGAACATTTGAT 2329  
 Db 802 CTTCAAGCTCGGGTTAAAGTTTCCAGTATAGAGAGACTGTCTACTAGAACATTTGAT 861  
 QY 2330 TGAATTTTTCAGGTCTATTCAGATCTCTAGATGTTATTTTAAAGA 2375  
 Db 862 TGAATTTTTCAGGTCTATTCAGATCTCTAGATGTTATTTTAAAGA 907

# RESULT 8 CB243715/c

LOCUS CB243715/c 745 bp mRNA linear EST 12-FEB-2003  
 DEFINITION UI-CF-FNO-9gf-d-02-0-UI.s1 UI-CF-FNO Homo sapiens cDNA clone  
 UI-CF-FNO-9gf-d-02-0-UI 3', mRNA sequence.

ACCESSION CB243715  
 VERSION CB243715.1 GI:28365359

KEYWORDS EST.

SOURCE Homo sapiens (human)

## ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 745)

Bonaldo,M.F., Lennon,G. and Soares,M.B

Normalisation and subtraction: two approaches to facilitate gene

## JOURNAL

Genome Res. 6 (9), 791-806 (1996)

## MEDLINE

97044477

## PUBMED

8889548

## COMMENT

Contact: McCray, PB

McCray Lab

University of Iowa

2024 University of

Tel: 319 356 4866

Fax: 319 356 7171

Email: paul-mccray@uiowa.edu

Tissue Procurement: Dr. M. J. Welsh, University of Iowa

CDNA Library preparation: Dr. M. Bento Soares, University of Iowa

CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Researchers may obtain clones from Research

Genetics (www.resgen.com) or from Open Biosystems

(www.openbiosystems.com).

Seq primer: M13 FORWARD

POLYA=Yes.

Location/Qualifiers

1..745

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="UI-CF-FNO-9gf-d-02-0-UI"

/tissue\_type="Human Lung Epithelial cells"

/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"

/clone\_lib="UI-CF-FNO"

/notes="Organ: Lung; Vector: p7T3-Pac (Pharmacia) with a

modified polylinker; Site 1: EcoR I; Site 2: Not I;

UI-CF-FNO is a subtracted cDNA library derived from two

normalized Human lung epithelial cell libraries (EN1 and

DUI). The library was subtracted according to according to

Bonaldo, Lennon and Soares, Genome Research, 6:791-806,

1996. For additional information, contact:

bento-soares@uiowa.edu

TAG LIB=UI-CF-FNO

TAG\_TISSUE=Human Lung Epithelial Cell Lines untreated LPS

```
6hr to LPS 24h
TAG SEQ=CTGCTCAGGT"
BASE COUNT 249 a 132 c 115 g 249 t
ORIGIN
Query Match 25.9%; Score 738.2; DB 14; Length 745;
Best Local Similarity 98.9%; Pred. No. 2.8e-71;
Matches 737; Conservative 5; Mismatches 3; Indels 0; Gaps 0;
Qy 2088 AGAAATTTAAACCTTATAATATTTTCTATCAARAWAWATTTTACAGTCGTCCTTT 2147
Db 745 AGAAATTTAAACCTTATAATATTTTCTATCAAGATGTATTTTACAGTCGTCCTTT 686
Qy 2148 TACTCTCATTTAGCAAAAAGATAAGATTTTATTTTGGTATTTTACAGAAATTCACAGTA 2207
Db 685 TACTCTCATTTAGCAAAAAGATAAGATTTTATTTTGGTATTTTACAGAAATTCACAGTA 626
Qy 2308 CGAAGATATCTGCGATGGTGGAAATCAGGTTCAAGCAACGCTTCTCANTTAAGTATA 2267
Db 625 CGAAGATATCTGCGATGGTGGAAATCAGGTTCAAGCAACGCTTCTCANTTAAGTATA 566
Qy 2268 TACCTCAGCTGCGGGTTAAAGTTTCCAGATATAGAGAGACTGCTCACTAGGAACATGT 2327
Db 565 TACCTCAGCTGCGGGTTAAAGTTTCCAGATATAGAGAGACTGCTCACTAGGAACATGT 506
Qy 2328 ATTGATTTATTCAGGTCATTTAGATCTTCTAGATGTATTTTAAAGAAATGCTTTTGGT 2387
Db 505 ATTGATTTATTCAGGTCATTTAGATCTTCTAGATGTATTTTAAAGAAATGCTTTTGGT 446
Qy 2388 TATGTGTTGTACACAGTTTAAACACTCCATAATGTTTATGTCAGCCAAAGAGGACTAAC 2447
Db 445 TATGTGTTGTACACAGTTTAAACACTCCATAATGTTTATGTCAGCCAAAGAGGACTAAC 386
Qy 2448 AAAGCTGAATCTCAGAGAAACAATTTGCTTTACTAAGCTGAGTCAACTTGAGACGGAAT 2507
Db 385 AAAGCTGAATCTCAGAGAAACAATTTGCTTTACTAAGCTGAGTCAACTTGAGACGGAAT 326
Qy 2508 TCTAACAATCGGCACCTGTAGTGGTGGTCTTACCACATATGACTTTTAAACATGTTTA 2567
Db 325 TCTAACAATCGGCACCTGTAGTGGTGGTCTTACCACATATGACTTTTAAACATGTTTA 266
Qy 2568 TATCATTTTAAATTTTATGATGAGTGTGTCAGGAGAAATGTAATGTTCTATATGAA 2627
Db 265 TATCATTTTAAATTTTATGATGAGTGTGTCAGGAGAAATGTAATGTTCTATATGAA 206
Qy 2628 ATTCTTTTCAAGTTTGTTCATTAATAACAGTTTATTAATTAATCAAGCTTATAGATTT 2687
Db 205 ATTCTTTTCAAGTTTGTTCATTAATAACAGTTTATTAATTAATCAAGCTTATAGATTT 146
Qy 2688 GTGCTGCTGCAACTGCTGTGAAAATTTCTCTGAGTAAATTCCTGATTAATTCGATGAT 2747
Db 145 GTGCTGCTGCAACTGCTGTGAAAATTTCTCTGAGTAAATTCCTGATTAATTCGATGAT 86
Qy 2748 ACCAACCCTGAGATTTTGTCAACCTGATTAAGTCAATATGAATGATTAAGATGTGAG 2807
Db 85 ACCAACCCTGAGATTTTGTCAACCTGATTAAGTCAATATGAATGATTAAGATGTGAG 26
Qy 2808 AACAAAAAAGAAAAAAGAAAAA 2832
Db 25 AACACTGAAAAAAGAAAAAAGAAAAA 1
RESULT 9
BM978800/c
LOCUS
DEFINITION
UI-CF-DUI-1-adp-g-07-0-UI.61 UI-CF-DUI Homo sapiens cDNA clone
UI-CF-DUI-1-adp-g-07-0-UI 3', mRNA sequence.
ACCESSION
BM978800
VERSION
BM978800.1 GI:19598602
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 745)  
Bonaldo, M.F., Lennon, G. and Soares, M.B.  
Normalization and subtraction: two approaches to facilitate gene  
discovery  
Genome Res. 6 (9), 791-806 (1996)  
97044477  
8889548  
Contact: McCray, PB  
McCray Lab  
University of Iowa  
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA  
Tel: 319 356 4866  
Fax: 319 356 7171  
Email: paul-mccray@uiowa.edu  
Tissue Procurement: Dr. M. J. Welsh, University of Iowa  
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research  
Genetics (www.resgen.com) or from Open Biosystems  
(www.openbiosystems.com)  
Seq primer: M13 FORWARD  
PolyA=Yes.

Location/Qualifiers  
1. 745  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-CF-DUI-1-adp-g-07-0-UI"  
/tissue\_type="Primary Lung Epithelial Cells"  
/dev\_stage="Adult"  
/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
/clone\_lib="UI-CF-DUI"  
/notes="Organ: Lung; Vector: p773-Pac (Pharmacia) with a  
modified polylinker; Site 1: EcoR I; Site 2: Not I;  
UI-CF-DUI is a normalized cDNA library containing the  
following tissue(s): Primary Lung Epithelial Cells The  
library was constructed according to Bonaldo, Lennon and  
Soares, Genome Research, 6:791-806, 1996. First strand  
cDNA synthesis was primed with an oligo-dT primer  
containing a Not I site. Double stranded cDNA was ligated  
to an EcoR I adaptor, digested with Not I, and cloned  
directionally into p773-Pac vector. The oligonucleotide  
used to prime the synthesis of first-strand cDNA contains  
a library tag sequence that is located between the Not I  
site and the (dT)18 tail. The sequence tag for this  
library is GGCTGTAGGC.  
TAG LIB=UI-CF-DUI  
TAG TISSUE=Lung Epithelial Cells Tissue nos 359-368  
TAG SEQ=GGCTGTAGGC"

BASE COUNT 248 a 132 c 117 g 248 t  
ORIGIN

Query Match 25.4%; Score 723; DB 12; Length 745;  
Best Local Similarity 98.5%; Pred. No. 1.2e-69;  
Matches 734; Conservative 5; Mismatches 5; Indels 1; Gaps 1;

Qy 2089 GAAAAATTAACCTTATAATATTTTCTAT-CAARAWAWATTTTACAGTCGTCCTTT 2147  
Db 745 GAAAAATTAACCTTATAATATTTTCTATCAAGATGTATTTTACAGTCGTCCTTT 686  
Qy 2148 TACTCTCATTTAGCAAAAAGATAAGATTTTATTTTGGTATTTTACAGAAATTCACAGTA 2207  
Db 685 TACTCTCATTTAGCAAAAAGATAAGATTTTATTTTGGTATTTTACAGAAATTCACAGTA 626  
Qy 2208 CGAAGATATCTGCGATGGTGGAAATCAGGTTCAAGCAACGCTTCTCANTTAAGTATA 2267  
Db 625 CGAAGATATCTGCGATGGTGGAAATCAGGTTCAAGCAACGCTTCTCANTTAAGTATA 566  
Qy 2268 TACCTCAGCTGCGGGTTAAAGTTTCCAGATATAGAGAGACTGTCACCTAGGAACATGT 2327  
Db 565 TACCTCAGCTGCGGGTTAAAGTTTCCAGATATAGAGAGACTGTCACCTAGGAACATGT 506



QY 2328 ATTGATTATTTCAGTTCATTTGAGATCTCTAGATGTTATTTTAAAAAGAAAGCTTTTGGT 2387  
Db 505 ATTGATTATTTCAGTTCATTTGAGATCTCTAGATGTTATTTTAAAAAGAAAGCTTTTGGT 446  
QY 2388 TATGTTGTTGTTACACAGTAACTAACTCATAATGTTTCATGTCAGCCAAAGAGGACTAAC 2447  
Db 445 TATGTTGTTGTTACACAGTAACTCATAATGTTTCATGTCAGCCAAAGAGGACTAAC 386  
QY 2448 AAAGCTGAATCTCAGAGAACAAATTTGCTTTTACTTAAGCTGAGTCAACTTGAGAGCGAACT 2507  
Db 385 AAAGCTGAATCTCAGAGAACAAATTTGCTTTTACTTAAGCTGAGTCAACTTGAGAGCGAACT 326  
QY 2508 TCTAACATCGGCACCTGTTAGTGGCTGTTCTTACCACTATGATTTTAAAAAGATGTTTA 2567  
Db 325 TCTAACATCGGCACCTGTTAGTGGCTGTTCTTACCACTATGATTTTAAAAAGATGTTTA 266  
QY 2568 TATCATTTTAAATTTTATGATACGGTAGTGTCTAGGAGAAATGTAATGTTCTATATGAA 2627  
Db 265 TATCATTTTAAATTTTATGATACGGTAGTGTCTAGGAGAAATGTAATGTTCTATATGAA 206  
QY 2628 ATTCTTTTCAAGTTTGTTCATTAATAACAGTTATTAAATTAATCAGCGTTAGAGTTT 2687  
Db 205 ATTCTTTTCAAGTTTGTTCATTAATAACAGTTATTAAATTAATCAGCGTTAGAGTTT 146  
QY 2688 GTGCTGTCGAACCTGCTGTGAAATTTCTCTGAGTAATTTCTGATTTCTGATGATCCAG 2747  
Db 145 GTGCTGTCGAACCTGCTGTGAAATTTCTCTGAGTAATTTCTGATTTCTGATGATCCAG 86  
QY 2748 ACCAACCTGAGATTTTGTCAACCTGATTAAGTCAATATGATTAATGATTAATAAGATGTGAG 2807  
Db 85 ACCAACCTGAGATTTTGTCAACCTGATTAAGTCAATATGATTAATGATTAATAAGATGTGAG 26  
QY 2808 AACAAAAAAGAAAAAAGAAAAA 2832  
Db 25 AACACTGAAAAAAGAAAAAAGAAAAA 1

RESULT 10  
BU854444  
LOCUS  
DEFINITION AGENCOURT\_10409743 NIH\_MGC\_82 Homo sapiens cDNA clone IMAGE:6621725  
5' mRNA sequence.  
ACCESSION BU854444  
VERSION BU854444.1 GI:24039410  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 782)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaapb@mail.nih.gov  
Tissue Procurement: CLONTECH  
CDNA Library Preparation: CLONTECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LCM2873 row: o column: 05  
High quality sequence stop: 634.  
Location/Qualifiers  
1..782  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6621725"  
/lab\_host="DH10B (T1 phage-resistant)"  
/clone\_lib="NIH\_MGC\_82"

FEATURES  
Source

/note="Organ: testis; Vector: pDNR-LIB (Clontech); Site\_1:  
SfiI (ggcgctggcc); Site\_2: SfiI (ggcattatggcc); 5' and  
3' adaptors were used in cloning as follows: 5' adaptor  
sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor sequence:  
5'-ATTCTAGAGCCGCGCCGACATG-dt(30)BN-3' (where B = A,  
C, or G and N = A, C, G, or T). Average insert size  
1.35 kb (range 0.9-4.0 kb). 14/15 colonies contained  
inserts by PCR. This library was enriched for full-length  
clones and was constructed by Clontech Laboratories (Palo  
Alto, CA)."  
BASE COUNT 244 a 123 c 142 g 271 t 2 others  
ORIGIN  
Query Match 25.2%; Score 718.6; DB 13; Length 782;  
Best Local Similarity 98.5%; Pred. No. 3.6e-69;  
Matches 729; Conservative 5; Mismatches 5; Indels 1; Gaps 1;  
QY 1899 GCAGTGACCAAGAACCCACCAAACTAGGCTGCTGTTGTAAGAGGCAATCATTTTGC 1958  
Db 3 GCAGTGACCAAGAACCCACCAAACTAGGCTGCTGTTGTAAGAGGCAATCATTTTGC 62  
QY 1959 CATTTGTGAAGTTGTTGGATTAGTAAATGTAAGTAAAGCTTTGTAATTTTGA 2018  
Db 63 CATTTGTGAAGTTGTTGGATTAGTAAATGTAAGTAAAGCTTTGTAATTTTGA 122  
QY 2019 GAACTTTTAAATGTTCCAAATACCTATTTTCAAAGGGTAATCGTAAGATGTTAACCC 2078  
Db 123 GAACTTTTAAATGTTCCAAATACCTATTTTCAAAGGGTAATCGTAAGATGTTAACCC 182  
QY 2079 TTGGTATTAGAAAAATTAACCTTATATATTTTCTATCAARAWMAWATTTTACAGT 2138  
Db 183 TTGGTATTAGAAAAATTAACCTTATATATTTTCTATCAARAWMAWATTTTACAGT 242  
QY 2139 CGTGCCTTTTACTCTCATTTAGCAAAAAAGATAAGATTTTATTTTGTGTTTCAAGAAT 2198  
Db 243 CGTGCCTTTTACTCTCATTTAGCAAAAAAGATAAGATTTTATTTTGTGTTTCAAGAAT 302  
QY 2199 TCCAGGTAGCAAGATATCTGATGGTGAAATCAGGTTCAAGCAACGTAATTTGCATT 2258  
Db 303 TCCAGGTAGCAAGATATCTGATGGTGAAATCAGGTTCAAGCAACGTAATTTGCATT 362  
QY 2259 AACTGATAATACCTCAGCTCGGGGTTAAAGTTTCCAGTATAGAGAGACTGTCACCTAG 2318  
Db 363 AACTGATAATACCTCAGCTCGGGGTTAAAGTTTCCAGTATAGAGAGACTGTCACCTAG 422  
QY 2319 GAACATTTGATTGATTTTATTTTCAAGTCACTGATGTTTATTTTAAAGAAATG 2378  
Db 423 GAACATTTGATTGATTTTATTTTCAAGTCACTGATGTTTATTTTAAAGAAATG 482  
QY 2379 CTTTTTGGTTATGTTGTTGTTACCACTTAACTCCATTAATGTTTCATGTCAGCCAAAGA 2438  
Db 483 CTTTTTGGTTATGTTGTTGTTACCACTTAACTCCATTAATGTTTCATGTCAGCCAAAGA 542  
QY 2439 GGACTAACCAAGCTGAAATCTCAGAGAACAAATTTGCTTTTACTAAGTCAGTCACTTGA 2498  
Db 543 GGACTAACCAAGCTGAAATCTCAGAGAACAAATTTGCTTTTACTAAGTCAGTCACTTGA 602  
QY 2499 GAGGAACTTCTAACAATGCGCACTAGTGGCTGTTCTACCACTATGATCTTTAAA 2558  
Db 603 GAGGAACTTCTAACAATGCGCACTAGTGGCTGTTCTACCACTATGATCTTTANA 662  
QY 2559 ACATGTTTATATCATTTTAAATTTTATGATACGGTAGTGTGAGGAGAAATGTAAT-GT 2617  
Db 663 ACATGTTTATATCATTTTAAATTTTATGATACGGTAGTGTGAGGAGAAATGTAATGTT 722  
QY 2618 TCTATATGAATTCCTTTT 2637  
Db 723 TCTATATGAATTCCTTTT 742  
RESULT 11  
BM977212/c  
LOCUS BM977212 linear 711 bp mRNA EST 21-FEB-2003



```

DEFINITION  UI-CF-DUI-ads-1-04-0-UI.s1 UI-CF-DUI Homo sapiens cDNA clone
VERSION      BM977212
KEYWORDS     BM977212.1 GI:19595389
SOURCE       Homo sapiens (human)
ORGANISM     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 711)
AUTHORS      Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE        Normalization and subtraction: two approaches to facilitate gene
              discovery
JOURNAL      Genome Res. 6 (9), 791-806 (1996)
MEDLINE      97044477
PUBMED       8889548
COMMENT      Contact: McCray, PB
              McCray Lab
              University of Iowa
              2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
              Tel: 319 356 4866
              Fax: 319 356 7171
              Email: paul-mccray@uiowa.edu
              Tissue Procurement: Dr. M. J. Welsh, University of Iowa
              cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
              cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
              DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
              Clone Distribution: Researchers may obtain clones from Research
              Genetics (www.resgen.com) or from Open Biosystems
              (www.openbiosystems.com).
              Seq primer: M13 FORWARD
              PolyA=yes.

FEATURES             Location/Qualifiers
     1..711
         /organism="Homo sapiens"
         /mol_type="mRNA"
         /db_xref="taxon:9606"
         /clone="UI-CF-DUI-ads-1-04-0-UI"
         /tissue_type="Primary Lung Epithelial Cells"
         /dev_stage="Adult"
         /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
         /clone_lib="UI-CF-DUI"
         /notes="Organ: Lung; Vector: p7T3-Pac (Pharmacia) with a
         modified polylinker; Site1: Ecor I; Site2: Not I;
         UI-CF-DUI is a normalized cDNA library containing the
         following tissue(s): Primary Lung Epithelial Cells The
         library was constructed according to Bonaldo, Lennon and
         Soares, Genome Research, 6:791-806, 1996. First strand
         cDNA synthesis was primed with an oligo-dT primer
         containing a Not I site. Double stranded cDNA was ligated
         to an Ecor I adaptor, digested with Not I, and cloned
         directionally into p7T3-Pac vector. The oligonucleotide
         used to prime the synthesis of first-strand cDNA contains
         a library tag sequence that is located between the Not I
         site and the (dT)18 tail. The sequence tag for this
         library is GGCTGTAGGC.
         TAG LIB=UI-CF-DUI
         TAG TISSUE=Lung Epithelial Cells Tissue nos 359-368
         TAG_SEQ=GGCTGTAGGC"
BASE COUNT      236 a 129 c 112 g 232 t 2 others
ORIGIN

Query Match      24.8%; Score 705.4; DB 12; Length 711;
Best Local Similarity 98.9%; Pred. No. 1e-67;
Matches 703; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Qy 2115 CTATCAARWAWATTTTACAGTCGGTCTTTTACTCTCATTTAGCAAAAAGATAAGA 2174
      |||||:::|
Db 711 CTATCAAGATGTATATTTTACAGTCGGTCTTTTACTCTCATTTAGCAAAAAGATAAGA 652
      |||||:::|

Qy 2175 TTTTATTTGGTATTTCACAGAAATCCAGGTACAGATATCTGCATGGTGAATCA 2234
      |||||:::|
Db 651 TTTTATTTGGTATTTCACAGAAATCCAGGTACAGATATCTGCATGGTGAATCA 592
      |||||:::|

```

```

Qy 2235 GGTTCAAGCAACGTACTTTTCATTAATGATGATTAATACCTCAGCTGCGGGTTAAAGTTTTC 2294
      |||||:::|
Db 591 GGTTCAAGCAACGTACTTTTCATTAATGATGATTAATACCTCAGCTGCGGGTTAAAGTTTTC 532
      |||||:::|

Qy 2295 CCAGTATAGAGAGACTGTCTACAGGACATGTTATGATTTATTCAGTTCATTCAGATCT 2354
      |||||:::|
Db 531 CCAGTATAGAGAGACTGTCTACAGGACATGTTATGATTTATTCAGTTCATTCAGATCT 472
      |||||:::|

Qy 2355 TCTAGATGTTATTTTAAAAAGAAATGCTTTTGGTTATGTTGTTGCTACACAGTTAACTC 2414
      |||||:::|
Db 471 TCTAGATGTTATTTTAAAAAGAAATGCTTTTGGTTATGTTGTTGCTACACAGTTAACTC 412
      |||||:::|

Qy 2415 CATAATGTTTCATGTCAGCCAAAGAGGACTAAACCAAGCTGAAATCTCAGAGAAACAATTTG 2474
      |||||:::|
Db 411 CATAATGTTTCATGTCAGCCAAAGAGGACTAAACCAAGCTGAAATCTCAGAGAAACAATTTG 352
      |||||:::|

Qy 2475 CTTTACTTAAGCTGAGTCAACTTCAGAGGCACTCTTAACCAATGCCGACCTGTAGTGGC 2534
      |||||:::|
Db 351 CTTTACTTAAGCTGAGTCAACTTCAGAGGCACTCTTAACCAATGCCGACCTGTAGTGGC 292
      |||||:::|

Qy 2535 TGGTTCTTACCACATGACATTTTAAAAACATGTTTATATCATTTTAAATTTTATGATACGGT 2594
      |||||:::|
Db 291 TGGTTCTTACCACATGACATTTTAAAAACATGTTTATATCATTTTAAATTTTATGATACGGT 232
      |||||:::|

Qy 2595 AGTGTACGGAGAAATGTAATGTTCTATATGAAATTCCTTTTCAAGTTTGTTCATTAAT 2654
      |||||:::|
Db 231 AGTGTACGGAGAAATGTAATGTTCTATATGAAATTCCTTTTCAAGTTTGTTCATTAAT 172
      |||||:::|

Qy 2655 AACAGTTATTAATTTAAATCAGCTGTAGAGTTTGTGCTGCTGCNACTGCTGTGAAAATTT 2714
      |||||:::|
Db 171 AACAGTTATTAATTTAAATCAGCTGTAGAGTTTGTGCTGCTGCNACTGCTGTGAAAATTT 112
      |||||:::|

Qy 2715 CTCTGAGTAATCTGTGATTTGTGAATGATCCAGACCAACCTCGAGATTTTGTCAACCTGA 2774
      |||||:::|
Db 111 CTCTGAGTAATCTGTGATTTGTGAATGATCCAGACCAACCTCGAGATTTTGTCAACCTGA 52
      |||||:::|

Qy 2775 TTAAGTCAATATGAATGATTTAAAAAGATGTGAGAACAAAAAAGAAAAA 2825
      |||||:::|
Db 51 TTAAGTCAATATGAATGATTTAAAAAGATGTGAGAACAAAAAAGAAAAA 1

RESULT 12
BM970195/c
LOCUS      BM970195      705 bp      mRNA      linear      EST 20-FEB-2003
DEFINITION  UI-CF-EC1-abv-1-07-0-UI.s1 UI-CF-EC1 Homo sapiens cDNA clone
              UI-CF-EC1-abv-1-07-0-UI 3', mRNA sequence.
ACCESSION   BM970195
VERSION     BM970195.1 GI:19587782
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 705)
AUTHORS      Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE        Normalization and subtraction: two approaches to facilitate gene
              discovery
JOURNAL      Genome Res. 6 (9), 791-806 (1996)
MEDLINE      97044477
PUBMED       8889548
COMMENT      Contact: McCray, PB
              McCray Lab
              University of Iowa
              2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
              Tel: 319 356 4866
              Fax: 319 356 7171
              Email: paul-mccray@uiowa.edu
              cDNA Library preparation: Dr. M. J. Welsh, University of Iowa
              cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
              DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
              Clone Distribution: Researchers may obtain clones from Research
              Genetics (www.resgen.com) or from Open Biosystems
              (www.openbiosystems.com).

```







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: December 4, 2003, 19:41:23 ; Search time 21 Seconds  
(without alignments)  
1170.600 Million cell updates/sec

Title: US-10-074-527-2  
Perfect score: 3124  
Sequence: 1 MWGTAARRCPRELRCREA.....LLRDCTNSDQKWFKRMFL 581

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.\*

- 1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1638	52.4	578	4	US-09-217-306B-2
2	1625.5	52.0	560	4	US-09-217-306B-10
3	1142.5	36.6	633	2	US-08-648-298-2
4	1100.5	35.2	559	2	US-08-967-508-9
5	1100.5	35.2	559	3	US-08-967-506-9
6	1100.5	35.2	559	5	PCT-US94-02552-9
7	1100	35.2	517	2	US-08-967-508-19
8	1100	35.2	517	3	US-08-967-506-19
9	1100	35.2	517	5	PCT-US94-02552-19
10	1023	32.7	639	3	US-09-376-856-2
11	1017	32.6	638	3	US-09-347-488-2
12	991.5	31.7	603	4	US-09-795-926-43
13	974.5	31.2	506	4	US-09-795-926-31
14	967	31.0	631	4	US-09-795-926-41
15	950	30.4	535	4	US-09-795-926-29
16	697.5	22.3	366	4	US-09-795-926-39
17	680.5	21.8	269	4	US-09-795-926-27
18	450.5	14.4	321	4	US-09-795-926-35
19	433.5	13.9	224	4	US-09-795-926-23
20	256.5	8.2	240	4	US-09-795-926-33
21	234.5	8.1	209	4	US-09-795-926-37
22	239.5	7.7	143	4	US-09-795-926-21
23	237.5	7.6	112	4	US-09-795-926-25
24	155	5.0	1206	4	US-09-252-991A-19632
25	145	4.7	702	4	US-09-437-277-1
26	131.5	4.2	965	4	US-09-437-277-3
27	124.5	4.0	942	2	US-08-468-812-4

28	124.5	4.0	492	2	US-08-468-812-7	Sequence 7, Appli
29	124.5	4.0	492	4	US-08-590-563-4	Sequence 4, Appli
30	124.5	4.0	492	4	US-08-590-563-7	Sequence 7, Appli
31	124.5	4.0	492	4	US-09-770-621-4	Sequence 4, Appli
32	124.5	4.0	492	4	US-09-770-621-7	Sequence 7, Appli
33	112.5	3.6	1618	1	US-07-853-913-4	Sequence 4, Appli
34	112.5	3.6	2860	2	US-08-826-267-2	Sequence 2, Appli
35	109.5	3.5	654	4	US-09-252-991A-22603	Sequence 22603, A
36	108.5	3.5	127	1	US-08-392-828C-39	Sequence 39, Appli
37	108.5	3.5	127	3	US-09-330-945-39	Sequence 39, Appli
38	108.5	3.5	491	2	US-08-468-812-8	Sequence 8, Appli
39	108.5	3.5	491	4	US-08-590-563-8	Sequence 8, Appli
40	108.5	3.5	491	4	US-09-770-621-8	Sequence 8, Appli
41	106	3.4	345	4	US-09-252-991A-26330	Sequence 26330, A
42	105.5	3.4	480	2	US-08-468-812-5	Sequence 5, Appli
43	105.5	3.4	480	4	US-08-590-563-5	Sequence 5, Appli
44	105.5	3.4	480	4	US-09-770-621-5	Sequence 5, Appli
45	104.5	3.3	612	4	US-09-252-991A-30072	Sequence 30072, A

## ALIGNMENTS

RESULT 1  
US-09-217-306B-2  
; Sequence 2, Application US/09217306B  
; Patent No. 6465220  
; GENERAL INFORMATION:  
; APPLICANT: Haasan, Helle  
; APPLICANT: Clausen, Henrik  
; APPLICANT: Bennett, Eric P.  
; TITLE OF INVENTION: Glycosylation Using Galnac-T4 Transferase  
; FILE REFERENCE: 8850\*1  
; CURRENT APPLICATION NUMBER: US/09/217,306B  
; CURRENT FILING DATE: 1998-12-21  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 2  
; LENGTH: 578  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-217-306B-2

Query Match 52.4%; Score 1638; DB 4; Length 578;  
Best Local Similarity 56.8%; Pred. No. 1.8e-159;  
Matches 324; Conservative 78; Mismatches 150; Indels 18; Gaps 8;

Qy	23	VLLALLALAG-----LGSVLRAORGAGAAEPPTPRPGR-----EPVMPRPVPA	72
Db	13	LLLAFLTVAYIFVELLVSTPHASAGAG-ARELGSRRLSDLQKNTEDLSRPLVKKPPADS	71
Qy	73	NALGARGEAVRLQOGHEELRQESVRLHGINIYLSDRISLHRLPERWNPCKEKKVDY	132
Db	72	RALGEGWAKSQNLNDELKQOEELRYAINIYLSDRISLHRIEDKRWYCKSQFNY	131
Qy	133	DNLPRTSVIIAFNEAWSTLLRTVYSVLETSPDILLEEVILVDYSDREHLKERLANELS	192
Db	132	RTLPTTSVIIAFNEAWSTLLRTIHSVLETSPAVLLKEILVDDLSRVVLTQLETYIS	191
Qy	193	GLPKVRLIRANKKEGLVRLALGASARGDVLFTLCHCECHGEBPLLQRIHEESAV	252
Db	192	NLDRVRLIRANKKEGLVRLALGATGATGDLTFLYCHCECHGEBPLLIRIGRYETAV	251
Qy	253	VCVIDVIDWNTTEYLGNSGEPQIGGDFWELVTHVTWTPERIRIMQSPVDVIRSPTMAG	312
Db	252	VCVIDVIDWNTTEYLGNSGEPQIGGDFWELVTHVTWTPERIRIMQSPVDVIRSPTMAG	311
Qy	313	GLFAVSKKPYEYLGSDYDGTGMEVWGGNLEFSFRWOCGVLETHPCSHVGHVFPKQAPYS	372
Db	312	GLFAVSKKPYEYLGSDYDGTGMEVWGGNLEFSFRWOCGVLETHPCSHVGHVFPKQAPYA	371
Qy	373	RNKALANSVRAAEVWMDPEFKELYHHRNPRARLPFGDVTERTKQLRDKLQCKDFKWFLETY	432

Db 372 RPNFLQNTARAAEYVMDEYKEHFYRNPPARKEAYGDISERKLLRERLRCKSPDWYLYKNV 431  
Qy 433 YPELHVPEDEPRGFMQLQKGLTDYCFDYNPPDENQIVGHQVILYLCHGMQNOFFEYTS 492  
Db 432 FPNLHVPEDEPRGFMQLQKGLTDYCFDYNPPDENQIVGHQVILYLCHGMQNOFFEYTS 490  
Qy 493 QKEIRYNTHOPEGCIAVEAGMDTLIMHLCEE---TAPENQKFILOEDGSLFHEQSKCVQ 549  
Db 491 NKEIRFNS-VTELCAEVPQKQVYGMQNCPKDGPVPANIIWHFKEDGTIFHPSGLCLS 549  
Qy 550 AARKESSDSFVPLLRDCTNSD-HQKWFPE 578  
Db 550 AYRTPEGRPDVQ-MRTCDALDKNQIWSFEK 578

## RESULT 2

US-09-217-306B-10  
; Sequence 10, Application US/09217306B  
; Patent No. 6465220  
; GENERAL INFORMATION:  
; APPLICANT: Hassan, Helle  
; APPLICANT: Clausen, Henrik  
; APPLICANT: Bennett, Eric P.  
; TITLE OF INVENTION: Glycosylation Using Galnac-T4 Transferase  
; FILE REFERENCE: 8850\*1  
; CURRENT APPLICATION NUMBER: US/09/217,306B  
; CURRENT FILING DATE: 1998-12-21  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 10  
; LENGTH: 560  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Interferon fragment-Galnac-T4 predicted coding region  
; NAME/KEY: PEPTIDE  
; LOCATION: (1)..(25)  
; OTHER INFORMATION: human interferon signal peptide  
; NAME/KEY: MISC FEATURE  
; LOCATION: (37)..(560)  
; OTHER INFORMATION: Galnac-T4  
US-09-217-306B-10

Query Match 52.0%; Score 1625.5; DB 4; Length 560;  
Best Local Similarity 59.4%; Pred. No. 3.2e-156;  
Matches 309; Conservative 75; Mismatches 129; Indels 7; Gaps 5;

Qy 63 PVMRPPVPANALGARGEAVRLQGEELRLQEEVSRLHQLINYLSDRLSHRLPERWN 122  
Db 44 PLYKKPPADSRALGEWGRKSLQNEDELKQEEELIERYAINYLSDRLSHRLHIEKRM 103  
Qy 123 PLCKEKYDYNLPRTSVIIAFYNEAWSTLRTVYSLVLETSFDILLEEVILVDDYSDEH 182  
Db 104 YECKSQKFNRYLPTTSVIIAFYNEAWSTLRTVYSLVLETSFPAVLKEIILVDDLSRVY 163  
Qy 183 LKERLANELSLPKVRLIRANKREGLVRARLLGASARGDVLTFDCHCECHGWLPLLL 242  
Db 164 LKTOLETYISNLDRLVRLIRANKREGLVRARLLGATFATGDLVTLFLYCHCECNGLWLEPL 223  
Qy 243 ORIEEESAVCPVIDLWNTTFEVLGNSGEPQGGFDWRLVFTVHTVPERIRMQSPV 302  
Db 224 ERIGRYETAVCPVIDLWNTTFEYMQIGEPMLGGFDWRLVFTVHTVPERIRMQSPV 283  
Qy 303 DVIRSPMTMAGGLFAVSKKYFYLGSYDTGMVWGGENLEFSGFRWQCQGVILETHPCSHVG 362  
Db 284 DPIRSPMTMAGGLFAVSKKYFYLGSYDTGMVWGGENLEFSGFRWQCQGVILETHPCSHVG 343  
Qy 363 HVFPKQAPYSRKNALANSVRAAEVWMDPEFKELYHRNPRARLEFPDGVTERKQLRDLQOC 422  
Db 344 HVFPKQAPYRPNFLQNTARAAEYVMDEYKEHFYRNPPARKEAYGDISERKLLRERLR 403

Qy 423 KDFKWELETYVPELHVPEDEPRGFMQLQKGLTDYCFDYNPPDENQIVGHQVILYLCHGM 482  
Db 404 KSPDWYLYKNVFPNHLVPEDEPRGFMQLQKGLTDYCFDYNPPDENQIVGHQVILYLCHGM 462  
Qy 483 GQNOFFEYTSQKFIYNTHOPEGCIAVEAGMDTLIMHLCEE---TAPENQKFILOEDGSL 539  
Db 463 GQNOFFEYTSQKFIYNTHOPEGCIAVEAGMDTLIMHLCEE---TAPENQKFILOEDGSL 539  
Qy 540 FHEQSKKCVQAAKESSDSFVPLLRDCTNSD-HQKWFPE 578  
Db 522 FHEHSGLCLSAYRTPEGRPDVQ-MRTCDALDKNQIWSFEK 560

## RESULT 3

US-08-648-298-2  
; Sequence 2, Application US/08648298  
; Patent No. 5871990  
; GENERAL INFORMATION:  
; APPLICANT: Henrik Clausen  
; APPLICANT: Eric Paul Bennett  
; TITLE OF INVENTION: UDP-N-acetyl-alpha-D-galactosamine;polypeptide  
; TITLE OF INVENTION: N-acetylgalactosaminyltransferase Galnac-T3  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Darby & Darby PC  
; STREET: 805 Third Avenue  
; CITY: New York  
; STATE: NY  
; ZIP: 10022  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/648,298  
; FILING DATE: 15-JUN-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Green, Reza  
; REGISTRATION NUMBER: 39,475  
; REFERENCE/DOCKET NUMBER: 4035/08B65  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212527700  
; TELEFAX: 2127536237  
; TELEX: 236687  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 633 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; TISSUE TYPE: Submaxillary gland  
US-08-648-298-2

Query Match 36.6%; Score 1142.5; DB 2; Length 633;  
Best Local Similarity 44.9%; Pred. No. 1.9e-108;  
Matches 244; Conservative 90; Mismatches 163; Indels 47; Gaps 14;

Qy 62 EPVMPRPPVPANALGARGEAVR-LQGEELRLQEEVSRLHQLINYLSDRLSHRL-PE 119  
Db 108 KPVLDPPQSNAPGASGKAFKTNLSVEQKEKEGECFNAFASDRLSHRLDGLPD 167  
Qy 120 RWNPLCKEKYD-YDNLPRTSVIIAFYNEAWSTLRTVYSLVLETSFDILLEEVILVDDYS 178  
Db 168 TRPECEIQKFKRCPPLTTSVIIAFYNEAWSTLRTVYSLVLETSFDILLEEVILVDDYS 227  
Qy 179 DREHLKERLANELSLPKVRLIRANKREGLVRARLLGASARGDVLTFDCHCECHGWL 238  
Db 228 VDEYHLKDLDEYKQKFSIVKIVQRERKGLITARLLGATVATSETLTFDHAHCFCFYGL 287



```

; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/602,830
; FILING DATE: 13 No. 6096512ember 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Darnley Jr., James D.
; REGISTRATION NUMBER: 33,673
; REFERENCE/DOCKET NUMBER: 4755.P CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 616-833-2210
; TELEFAX: 616-833-8897
; TELEX: 224401
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 559 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-967-506-9

Query Match 35.2%; Score 1100.5; DB 3; Length 559;
Best Local Similarity 43.0%; Pred. No. 3.3e-104;
Matches 240; Conservative 86; Mismatches 171; Indels 61; Gaps 17;

Qy 38 RAQRGAGAGAAEPGPRTPRGRPEVMPRPVPANALGARGAVRLQLOGEELRLQEEES 97
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
37 KKEGLPAGDV-----LEPVQK-----PHEGPGMGKPVVIPKEDQE--RKEM 78
Qy 98 VRLHQINILSDRISLHRRLPERNPLCKEKKYDYNLPTSVIIAFVNEAWSTLLRTVY 157
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
79 FKINQPNLMASEMIALNRSPLDVLEGCKTKVYP-DNLPTTSVVIVFHNESWSTLLRTVH 137
Qy 158 SVLETSPIILLEEVILVDYSDRHLKERLANELSGLP-KVRLIRANKREGLVRARLGA 216
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
138 SVINRSPRHLMEILVDDASERDFLKRPLESYVKLKVPHVIRMEQSGILIRALKGA 197
Qy 217 SAARGDVLTLFDCHEGHEGHEGHEGHEGHEGHEGHEGHEGHEGHEGHEGHEGHEG 276
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
198 AVSGQVITFLDAHCECTVGWLEPPLARIKHDRKTVVCPDIIDVISDDTFFYMAGS-DMTY 256
Qy 277 GGFDRWLVTWHTVPERIRMQSPVDV-IRSPMTAGGLFAVSKKYFEYLGSDYTGMEVW 335
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
257 GGFNWKLNFRVYVPQREMDRKGDRTLPVTPMTAGGLFSIDRDYFQEGTIDYDAGMDIW 316
Qy 336 GGENLEFSFRWQCGGVLETHPCSHVGHVFPKQAPYS-----RNKALANSVRAAEVW 387
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
317 GGENLEISFRWQCGGLEIVTCSHVGHVFRKATPYTPPGGTGQIINK---NNRRLAEVW 373
Qy 388 MDEPKELYHHRNPRARLEPFGDVTERRKQDKCKQDKFVPLETVYPDELHVPEDRPGFFG 447
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
374 MDEPKNFYIISPGVTKVYDYGDISRSLRGLRHKLQCRPFSWYLENIYDPSQIPRH---YFS 430
Qy 448 MLQNKGL-TDYCFDYNPDENQIVGHQVILYLCHGMGNQOFEYTSOKERYNTHOPEGC 506
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
431 LGEIRNVETNOCLDNMARKENKVG-----IFNCHGMGNQVFSYANKERTD----DLC 482
Qy 507 IAVBAGMDTLIMHICEETAPENQ-----KFILQEDGSLPHEOSKKCVQARKESSDSF 559
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
483 LDVSKLGNPVTMLKCHH-LKGNQLWEVDPVKLTILQ-----HVNNSQCLDKATDE--DSQ 533
Qy 560 VPLLRDCTNSDHQKFFK 577
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
534 VPSIRDCSGRSQQWLLR 551

```

## RESULT 6

PCT-US94-02552-9

```

; Sequence 9, Application PC/TUS9402552
; GENERAL INFORMATION:
; APPLICANT: Elhammer, Ake P.
; APPLICANT: Homa, Fred L.

```

```

; TITLE OF INVENTION: A Cloned DNA Encoding a UDP-GalNac:
; TITLE OF INVENTION: Polypeptide, N-Acetylgalactosaminyltransferase
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Upjohn Company, Corp. Intellectual
; ADDRESSEE: Property Law
; STREET: 301 Henrietta Street
; CITY: Kalamazoo
; STATE: Michigan
; COUNTRY: USA
; ZIP: 49001
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/02552
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Darnley Jr., James D.
; REGISTRATION NUMBER: 33,673
; REFERENCE/DOCKET NUMBER: 4755.P CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 616-385-5210
; TELEFAX: 616-385-6897
; TELEX: 224401
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 559 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US94-02552-9

```

```

Query Match 35.2%; Score 1100.5; DB 5; Length 559;
Best Local Similarity 43.0%; Pred. No. 3.3e-104;
Matches 240; Conservative 86; Mismatches 171; Indels 61; Gaps 17;

Qy 38 RAQRGAGAGAAEPGPRTPRGRPEVMPRPVPANALGARGAVRLQLOGEELRLQEEES 97
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
37 KKEGLPAGDV-----LEPVQK-----PHEGPGMGKPVVIPKEDQE--RKEM 78
Qy 98 VRLHQINILSDRISLHRRLPERNPLCKEKKYDYNLPTSVIIAFVNEAWSTLLRTVY 157
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
79 FKINQPNLMASEMIALNRSPLDVLEGCKTKVYP-DNLPTTSVVIVFHNESWSTLLRTVH 137
Qy 158 SVLETSPIILLEEVILVDYSDRHLKERLANELSGLP-KVRLIRANKREGLVRARLGA 216
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
138 SVINRSPRHLMEILVDDASERDFLKRPLESYVKLKVPHVIRMEQSGILIRALKGA 197
Qy 217 SAARGDVLTLFDCHEGHEGHEGHEGHEGHEGHEGHEGHEGHEGHEGHEGHEGHEG 276
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
198 AVSGQVITFLDAHCECTVGWLEPPLARIKHDRKTVVCPDIIDVISDDTFFYMAGS-DMTY 256
Qy 277 GGFDRWLVTWHTVPERIRMQSPVDV-IRSPMTAGGLFAVSKKYFEYLGSDYTGMEVW 335
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
257 GGFNWKLNFRVYVPQREMDRKGDRTLPVTPMTAGGLFSIDRDYFQEGTIDYDAGMDIW 316
Qy 336 GGENLEFSFRWQCGGVLETHPCSHVGHVFPKQAPYS-----RNKALANSVRAAEVW 387
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
317 GGENLEISFRWQCGGLEIVTCSHVGHVFRKATPYTPPGGTGQIINK---NNRRLAEVW 373
Qy 388 MDEPKELYHHRNPRARLEPFGDVTERRKQDKCKQDKFVPLETVYPDELHVPEDRPGFFG 447
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
374 MDEPKNFYIISPGVTKVYDYGDISRSLRGLRHKLQCRPFSWYLENIYDPSQIPRH---YFS 430
Qy 448 MLQNKGL-TDYCFDYNPDENQIVGHQVILYLCHGMGNQOFEYTSOKERYNTHOPEGC 506
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
431 LGEIRNVETNOCLDNMARKENKVG-----IFNCHGMGNQVFSYANKERTD----DLC 482

```



QY 507 IAVEAGMDTLIMHLCETAPENQ-----KFILQEDGSLFHEQSKKCVQARKESSDSF 559  
Db 483 LDVSKLNGPVTMLKCHH-LKGNQLWEYDPVKLTQ-----HVNQNCLDKATDE--DSQ 533  
QY 560 VPLLRDCTNSDHQKWFVK 577  
Db 534 VPSIRDCSGRSQQWLLR 551

RESULT 7  
US-08-967-508-19  
; Sequence 19, Application US/08967508  
; Patent No. 5910570  
; GENERAL INFORMATION:  
; APPLICANT: The Upjohn Company  
; APPLICANT: FOR U.S. PURPOSES ONLY: Elhammer, Ake P. and Homa, Fred L.  
; TITLE OF INVENTION: A Cloned DNA Encoding a UDP-GalNAC:  
; TITLE OF INVENTION: Polypeptide, N-Acetylgalactosaminyltransferase  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pharmacia & Upjohn Company, Intellectual  
; ADDRESSEE: Property Legal Services  
; STREET: 301 Henrietta Street  
; CITY: Kalamazoo  
; STATE: Michigan  
; COUNTRY: USA  
; ZIP: 49001  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/967,508  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/602,830  
; FILING DATE: 13 No. 5910570ember 1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Darnley Jr., James D.  
; REGISTRATION NUMBER: 33,673  
; REFERENCE/DOCKET NUMBER: 4755.P CP  
; TELEPHONE: 616-833-2210  
; TELEFAX: 616-833-8897  
; TELEX: 224401  
; INFORMATION FOR SEQ ID NO: 19:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 517 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-967-508-19

Query Match 35.2%; Score 1100; DB 2; Length 517;  
Best Local Similarity 44.1%; Pred. No. 3.3e-104;  
Matches 238; Conservative 84; Mismatches 168; Indels 50; Gaps 17;

QY 56 PRGRRPVPMPVPANALGARGEAVRLQQLQGEELRLQBSVRLHQINIYLSRISLHR 115  
Db 2 PSPA-LEPVQK---PHEGPGEMKPVVVPKEDQE--KMKEMFKINQFNLMASEMIALNR 54  
QY 116 RLPERWNPCKEKKYDYNLPRTSVIITAFYNEANSTLLRTVYSVLETSPIILLBEVLVD 175  
Db 55 SLDPVRLGCKTKYVP-DNLPTTSVIVFHNAAVSTLLRTVHVSINSPRHMLEEVLVD 113  
QY 176 DYSREHLKERLANELGLP-KVRLIRANKREGVLRALLGASAARGDVLTLFLDCHCECH 234  
Db 114 DASERDFLKPLESYVKLKVPHVHVRWEQSGRLIRALKGAAVSKGVITFLDAHCECT 173  
QY 235 EGMLEPLLRHIEEESAVVCPVIDIDWNTPEYLGNSGEPQIGGFDWRLVFTWHTVPERE 294

Db 174 VGMLEPLLARIKHDRKTVWCPDIIDVSDDTFEYNMAGS-DMTYGGFNKLNFRWYVPVQRE 232  
QY 295 RIRMQSPDV-IRSPTMAGGLFAVKYFVGLSGYDTGMVGVGGENLEFSPRIWQCGVL 353  
Db 233 MDRKGDRTLPVRTPTWAGGLFSDRYFOBIGYDAGMDIWGGENLEISPRIWQCGTLL 292  
QY 354 ETHPCSHVGHVFPKQAPYS-----RNKALANSVRAAEVWMDPEFKELYHYHRNPRARLE 405  
Db 293 EIVTCSHVGHVFRKATPTPTPGGTGQIINK--NNRRLAEVWMDPEKFNFFVILSPGVTKV 349  
QY 406 PFGDVTERRKQIRDLCKQDFKWFLETYVPELHVPEDRPGFPGMLQNLGL-TDYCFDYNPP 464  
Db 350 DYGDISSRLGLRHKLCRPFWSYLENIYDPSQIPRH---YFSLGEIRNVETNQCLDNMAR 406  
QY 465 DENQIVGHQVILYLCHGMGNORPEYTSQKEIRYNTHQPEGCIAVEAGMDTLIMHLCBEET 524  
Db 407 KENEKVG---IFNCHGMGNQVFSYNTANKEIRTD----DLCLDVSXLNGVPTWMLKCHH- 457  
QY 525 APENQ-----KFILQEDGSLFHEQSKKCVQARKESSDSFVPLLRDCTNSDHQKWFVK 577  
Db 458 LKGNQLWEYDPVKLTQ-----HVNQNCLDKATDE--DSQVPSIRDCSGRSQQWLLR 509

RESULT 8  
US-08-967-506-19  
; Sequence 19, Application US/08967506  
; Patent No. 6096512  
; GENERAL INFORMATION:  
; APPLICANT: The Upjohn Company  
; APPLICANT: FOR U.S. PURPOSES ONLY: Elhammer, Ake P. and Homa, Fred L.  
; TITLE OF INVENTION: A Cloned DNA Encoding a UDP-GalNAC:  
; TITLE OF INVENTION: Polypeptide, N-Acetylgalactosaminyltransferase  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pharmacia & Upjohn Company, Intellectual  
; ADDRESSEE: Property Legal Services  
; STREET: 301 Henrietta Street  
; CITY: Kalamazoo  
; STATE: Michigan  
; COUNTRY: USA  
; ZIP: 49001  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/967,506  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/602,830  
; FILING DATE: 13 No. 6096512ember 1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Darnley Jr., James D.  
; REGISTRATION NUMBER: 33,673  
; REFERENCE/DOCKET NUMBER: 4755.P CP  
; TELEPHONE: 616-833-2210  
; TELEFAX: 616-833-8897  
; TELEX: 224401  
; INFORMATION FOR SEQ ID NO: 19:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 517 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-967-506-19

Query Match 35.2%; Score 1100; DB 3; Length 517;  
Best Local Similarity 44.1%; Pred. No. 3.3e-104;

Matches	238;	Conservative	84;	Mismatches	168;	Indels	50;	Gaps	17;
QY	56	PRGRREPVRPPVPANALGARGEARLQIQGEELRLQBEESVRLHQINLYLSDRISLHR	115						
Db	2	PSA-LEPVOK- ---PHEGFGEMGKPVVYPKEDE--KMKEMPKINQFNLMASEIALNR	54						
QY	116	RLPERNPNLCKEKKYVDNLPRTSVIIAFYNEAWSTLLRTVYVSLTSPDILLBEVILVD	175						
Db	55	SLPDVRLGEGCKTKYP-DNLPTTSVIVFHNEAWSTLLRTVHVSINRPHRMLSEIIVLD	113						
QY	176	DYSDREHLKERLANELSGLP-KVRLIRANKREGIVRLARLIGASAARGDVLTFDLCHCECH	234						
Db	114	DASERDFKRLPLSEYVKJGVPHVIRMEQRSLIRARLKGAAVSKQGVITFLDAHCECT	173						
QY	235	EGMLEPQLQRIHEEASAVCPVIDIVDWNTFEYLGNSGEPQIGGDFWRLVFTWHTVPERE	294						
Db	174	VGMLEPLLARIKHDKTVVCPIDIVISDDTFEYWAGS-DMTYGGFNKWLFRWTFVPQRE	232						
QY	295	RIRMQSPVDV-IRSPPTMAGGLFAVSKKYFEYLGSYDTGMVWGGENLEFSFRIWQCQGVL	353						
Db	233	MDRRKGRDRTLVPRTPTMAGGLFSIDRDYFQEIQGYDAGMDIWGGENLEISFRIWQCQGV	292						
QY	354	ETHPCHSHGVHVPKQAPYS-----RNKALANSVRAAEVWMDPEFELYTHRNPRARLE	405						
Db	293	EIVTCSHGVHVFVKATPTTPGGTGTQIINK---NNRRLAEVWMDPEFNFFYIISPGVTKV	349						
QY	406	PFGDVTERKOLRDKLOCKDFKWFLETVYPBELHVPEDRPGFPGMLQNKGL-TDYCFDYNPP	464						
Db	350	DYGDISRLGLRHKLQCRPFSWYLENTYPPSQIPRH---YFSLGEIRNVETNQCLDNMAR	406						
QY	465	DENQIVGHQVITLYLCHGMGNQFPFYTSQKEIRYNTHQPECIAVEAGMDTLIMHLCEET	524						
Db	407	KEHEKVG----IFNCHGMGNQVFSYATNKEIRTD----DLCLDVLKNGPVTMLKCHH-	457						
QY	525	APENQ-----KFILQEDGSLFHEQSKCVQAARKSESSDFVPLLRDCTNSDHQKFFK	577						
Db	458	LKGNOLWEYDPVKLTLO-----HWNSNOCLDKATDE--DSOVPSIRDCSGRSOOWLLR	509						

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999

```

```

; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 517 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US94-02552-19

Query Match          35.2%; Score 1100; DB 5; Length 517;
Best Local Similarity 44.1%; Pred. No. 3.3e-104;
Matches 238; Conservative 84; Mismatches 168; Indels 50; Gaps 17.

QY      56 PRGRREPPWMPRPVPANALGARGEAVRLQLQGEELRLQOESVRLHQINLYLSDRISLHR 115
      |||      |||      |||      |||      |||      |||      |||      |||      |||
Db      2 P89A-LEPVQK----PHEGFGEMGKVPVPKEDQE--KMKEMPKINQFNLMASEMIALNR 54

QY      116 RLPERNPLCKEKYDYDNLPRTSVLIIFNEAWSLTLLRTVYSLVETSPDILLEEVILVD 175
      |||      |||      |||      |||      |||      |||      |||      |||      |||
Db      55 SLPDVELEGCKTVYP-DNLPTTSVVIVFNEAWSLTLLRTVHVSINRSPHMLIEEIVLVD 113

QY      176 DYSDRHLKRLANELSLGP-KVRLIRANKREGLVRLARLIGASAAAGDVLTLFDLCHCEH 234
      |||      |||      |||      |||      |||      |||      |||      |||      |||
Db      114 DASERDFLRPLESYVKLKVPHVIRMEQRSLIRLKGAAVSKGVITFLDAHCECT 173

QY      235 EGMLEPLLORIHEESAVCPVIDVDWNTFFYGLNLSGEPQIGCFDWRLLVFTVHTVPERE 294
      |||      |||      |||      |||      |||      |||      |||      |||      |||
Db      174 VGMLEPLLARIKHDRITVVCPIIDVTSDDTFFEYMAGS-DMTYGGFNKMLNFRWYPVPQRE 232

QY      295 RIRMQSPDVV-IRSPTMAGLPAVSKKYFEYLGSDYDTGMVWNGENLEPSFR1WQCGVL 353
      |||      |||      |||      |||      |||      |||      |||      |||      |||
Db      233 MDRKRGDRLPVKPTPTMAGLPSIDRDYFQEITGYDAGMDIWGEMLEISFR1WQCGGTL 292

QY      354 ETHPCSHGVHVFPKQAPYS-----RNKALANSVRAAEVWMDPEFKELYHHRNPRARLE 405
      |||      |||      |||      |||      |||      |||      |||      |||      |||
Db      293 EIVTCSHGVHVFPRKATPYTFPGGTGQIINK-----NNRRLAEVWMDPEKQVFIISPGVTKV 349

QY      406 PFGDVTTERQLRDLKLOCKDFKWFLETVYDELHVPEDRPGFFGLMQLNKL-TDYCFDYNPP 464
      |||      |||      |||      |||      |||      |||      |||      |||      |||
Db      350 DYGDISRLSLRLRHKLQCRPFSWYLENIYPSDSQ1PRH---YFSIGEIRNVTNOCQDNMAR 406

QY      465 DENOIVGHGVILYLCHGMGNQGFPEYTSOKERIYNTHQPEGCCIAVEAGMDTLMLHCEET 524
      |||      |||      |||      |||      |||      |||      |||      |||      |||
Db      407 KENEKVG----IFNCHGMGNQVFSYTANKEIRTD-----DLCLDVKLNGPVTMLKCHH- 457

QY      525 APENQ-----KFIIQEDGSLPHEOSKCKVCQVQAAKESDVSFVLIRDCNTSDHQKWFKK 577
      |||      |||      |||      |||      |||      |||      |||      |||      |||
Db      458 LKGNOLWEYDPVKTLQ-----HVNNSNOCCLDKATDE--DSOVPSIRDCSCGSQOOWLLR 509

```

```

RESULT 10
US-09-376-856-2
; Sequence 2, Application US/09376856
; Patent No. 6235510
; GENERAL INFORMATION:
; APPLICANT: KUMAR, SANJAY
; APPLICANT: VAN HORN, MARION M.
; APPLICANT: LARK, MICHAEL
; TITLE OF INVENTION: ppgAntase-T6
; FILE REFERENCE: GP-70638
; CURRENT APPLICATION NUMBER: US/09/376,856
; CURRENT FILING DATE: 1999-08-18
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 639
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
US-09-376-856-2

```

Query Match 32.7%; Score 1023; DB 3; Length 639;  
Best Local Similarity 42.4%; Pred. No. 3.9e-96;  
Matches 233; Conservative 78; Mismatches 203; Indels 3

Qy 48 AEPGPPPTPRGR-----EPVMPRPVPANALGA--RGEAVRLQLOGEELRQES 97  
Db 100 ALPOARRNQSGRRGGYRLIKQRRQDKEAPKRDWGADEGEV-----SEEEELTPFS 153  
Qy 98 VRLHQINILSDRISLHRLPERWNPCKEKKYDYNLPRTSVLIAPYNEAWSTLLRTVY 157  
Db 154 LDPGRQEAUSARIPLQALPEVRHPLCLQ-QHPQDSLPTASVILCFHDEAWSTLLRTVH 212  
Qy 158 SVLETSPIILLEEVILVDDYSREHLKERLANELSGLPKVLIRANKREGVLARLILGAS 217  
Db 213 SILDTVPRAFLKEIILVDDLSQQQQLKSALSEYVARLEGVKLRSNKRSLAIRARMLGAT 272  
Qy 218 AARGDVLTFDCHCEHGWLEPLLQRIHEESAVCPVIDVIDWNTFEYLGNSEGEPOIG 277  
Db 273 RATGDVLVFMFMDAHCCEHPGWLPLELLSRIAGRSRVSPVIDVIDWKTFOYY-PSKDLQRG 331  
Qy 278 GFDWRLVFTWHTTTPERIRMQSPVDVIRSPMTAGGLFAVSKYKFEYVLSGYDTGMEVWGG 337  
Db 332 VLDWKLDWFHEPPELPEHVRKALQSPISPIRSPVPGEVVAMDRHYFQNTGAYDSLMSURGG 391  
Qy 338 ENLEFSFRIWQCGVLETHPCSHVGHVFPKQ---APYSRNKALANSVRAAEVWMDPEKEL 394  
Db 392 ENLEFSFKAWLGGSVLEILPCSRVGHYQNDQSHSPDQEAATLRNRVRIAEATWLGSPKET 451  
Qy 395 YYHRNPRA----RLEPFGDVTERKQRLDKLQCKDFKWFLETVYPDELHVPEDRPGFFQMLQ 450  
Db 452 FYKHSPEAFSLSKAEK-PDCMERLQRLQRLGCRTHFWFLANVPYELYPSEPRPSFGKJH 510  
Qy 451 NKGLTDYCFDYNPDENQIVGHQVILYLCHGMQONOFFEYTSQKEIRYNTHQPEGCTAVE 510  
Db 511 NTGL-GLCADCQA--EGDILGCPMVLAPCSDRSQQQLQHTSRKEIHFGSPQ-HLCFAVR 566  
Qy 511 AGMDTLIMHLCBE--TAPENQKFIQEDGSLFHEQSKKCVQAAKESDSSFVPLLRDCTN 568  
Db 567 --QEOVILQNTCEGLAIHQHWDQFQENGMIHVLGSKCMEAVVQENNKDL--YLRPCDG 622  
Qy 569 SDHQKWFFKE 578  
Db 623 KARQWRFDQ 632

## RESULT 11

US-09-347-488-2  
; Sequence 2, Application US/09347488  
; Patent No. 6239266  
; GENERAL INFORMATION:  
; APPLICANT: Munroe, David  
; APPLICANT: Pribili, Ingrid  
; TITLE OF INVENTION: ZAP-3 TUMOR ASSOCIATED GENES AND THEIR  
; FILE REFERENCE: GSN-2PRV  
; CURRENT APPLICATION NUMBER: US/09/347,488  
; CURRENT FILING DATE: 1999-07-02  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 638  
; TYPE: PRT  
; ORGANISM: H. sapiens  
US-09-347-488-2

Query Match 32.6%; Score 1017; DB 3; Length 638;  
Best Local Similarity 42.2%; Pred. No. 1.6e-95;  
Matches 232; Conservative 78; Mismatches 204; Indels 36; Gaps 14;

Qy 48 AEPGPPPTPRGR-----EPVMPRPVPANALGA--RGEAVRLQLOGEELRQES 97  
Db 100 ALPOARRNQSGRRGGYRLIKQRRQDKEAPKRDWGADEGEV-----SEEEELTPFS 153  
Qy 98 VRLHQINILSDRISLHRLPERWNPCKEKKYDYNLPRTSVLIAPYNEAWSTLLRTVY 157  
Db 154 LDPGRQEAUSARIPLQALPEVRHPLCLQ-QHPQDSLPTASVILCFHDEAWSTLLRTVH 212

Qy 158 SVLETSPIILLEEVILVDDYSREHLKERLANELSGLPKVLIRANKREGVLARLILGAS 217  
Db 213 SILDTVPRAFLKEIILVDDLSQQQQLKSALSEYVARLEGVKLRSNKRSLAIRARMLGAT 272  
Qy 218 AARGDVLTFDCHCEHGWLEPLLQRIHEESAVCPVIDVIDWNTFEYLGNSEGEPOIG 277  
Db 273 RATGDVLVFMFMDAHCCEHPGWLPLELLSRIAGRSRVSPVIDVIDWKTFOYY-PSKDLQRG 331  
Qy 278 GFDWRLVFTWHTTTPERIRMQSPVDVIRSPMTAGGLFAVSKYKFEYVLSGYDTGMEVWGG 337  
Db 332 VLDWKLDWFHEPPELPEHVRKALQSPISPIRSPVPGEVVAMDRHYFQNTGAYDSLMSURGG 391  
Qy 338 ENLEFSFRIWQCGVLETHPCSHVGHVFPKQ---APYSRNKALANSVRAAEVWMDPEKEL 394  
Db 392 ENLEFSFKAWLGGSVLEILPCSRVGHYQNDQSHSPDQEAATLRNRVRIAEATWLGSPKET 451  
Qy 395 YYHRNPRA----RLEPFGDVTERKQRLDKLQCKDFKWFLETVYPDELHVPEDRPGFFQMLQ 450  
Db 452 FYKHSPEAFSLSKAEK-PDCMERLQRLQRLGCRTHFWFLANVPYELYPSEPRPSFGKJH 510  
Qy 451 NKGLTDYCFDYNPDENQIVGHQVILYLCHGMQONOFFEYTSQKEIRYNTHQPEGCTAVE 510  
Db 511 NTGL-GLCADCQA--EGDILGCPMVLAPCSDRSQQQLQHTSRKEIHFGSPQ-HLCFAVR 566  
Qy 511 AGMDTLIMHLCBE--TAPENQKFIQEDGSLFHEQSKKCVQAAKESDSSFVPLLRDCTN 568  
Db 567 --QEOVILQNTCEGLAIHQHWDQFQENGMIHVLGSKCMEAVVQENNKDL--YLRPCDG 622  
Qy 569 SDHQKWFFKE 578  
Db 623 KARQWRFDQ 632

## RESULT 12

US-09-795-926-43  
; Sequence 43, Application US/09795926  
; Patent No. 6555669  
; GENERAL INFORMATION:  
; APPLICANT: Donoho, Gregory  
; APPLICANT: Hilbun, Erin  
; APPLICANT: Turner, C. Alexander Jr.  
; APPLICANT: Friedrich, Glenn  
; APPLICANT: Abuin, Alejandro  
; APPLICANT: Zambrowicz, Brian  
; APPLICANT: Sands, Arthur T.  
; APPLICANT: Walke, D. Wade  
; APPLICANT: Wilganowski, Nathaniel L.  
; APPLICANT: Hu, Yi  
; APPLICANT: Kieke, James Alvin  
; APPLICANT: Potter, David George  
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
; FILE REFERENCE: POLYNUCLEOTIDES ENCODING THE SAME  
; FILE REFERENCE: LEX-0144-USA  
; CURRENT APPLICATION NUMBER: US/09/795,926  
; CURRENT FILING DATE: 2001-02-28  
; PRIOR APPLICATION NUMBER: US 60/185,920  
; PRIOR FILING DATE: 2000-02-29  
; PRIOR APPLICATION NUMBER: US 60/186,558  
; PRIOR FILING DATE: 2000-03-02  
; PRIOR APPLICATION NUMBER: US 60/191,849  
; PRIOR FILING DATE: 2000-03-24  
; NUMBER OF SEQ ID NOS: 47  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 43  
; LENGTH: 603  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-09-795-926-43

Query Match 31.7%; Score 991.5; DB 4; Length 603;  
Best Local Similarity 36.3%; Pred. No. 6.2e-93;  
Matches 231; Conservative 95; Mismatches 197; Indels 113; Gaps 19;

```
Qy 8 RRCPRRLRRGREALVLLALLALAGLSVLRAOR-----GAGAGAAEPGPPRTPRCRR 61
Db 2 RRKRRLQALVALVLAALLVLLPNVGLMALYRERQDGTGGSGHAVA-----48
Qy 62 EPVMPRPVPANALGARG-----EAVRLQLQ-----GEELR-----92
Db 49 -----PAAGQSHSRQKKTFFLGDGQKLDKWDHKEAIRDAQRVGNGEQGRPYPM 99
Qy 93 ---LQESVRLHQINIVLSDRISLHRRLPERWNPCKEKKYDYNLPRTSVILAFYNEAW 149
Db 100 DAERVQDAYRENGFNIVYSDKISLNSRLPDIRHPNCNKRY-LETLPNTSIIPFHNEGW 158
Qy 150 STLLRTVYVLETSPIILLEVILVDDYSDEHLEKRLANLGLPKVRLIRANKREGVL 209
Db 159 SLLRTVHSLNRPPELVAIVLVDPSDEHLEKRLANLGLPKVRLIRANKREGVL 218
Qy 210 RARLLGASARGDVLTFDCHCEHGLEPQLQRIHEESAVVCPVIDIDMNTTFYLG 269
Db 219 RTRMLGASVATGDTITFLDSHCEANVNWLPPLLDRIARNRKTIVCPMIDVIDHDDPRYET 278
Qy 270 NSGEPQGGFWRLVFTWHTVPERIRIMQSPVDVIRSPMTAGGLFAVSKKYFYLGSD 329
Db 279 QAGDAMRGAPDWEY--KRIPPELQKADPSDPFSPVMAGGLFAVDRKWFELGGYD 336
Qy 330 TGMVWGGENLEFSPRIWOCGGVLETHPCSHVGHVFPKQAPYS--RNKALA-NSVRAAEV 386
Db 337 PGLIEWGEQYEISFKVMWCGGRMEDIPCSRVGHIYRYKVPYKVPAGVSLARNLKRVAEV 396
Qy 387 WMDSPKELYHNRPRARLEPFGDVTERRKQLDKLQCKDFKWFLETV---YPELHVPEDRP 443
Db 397 WMDEVAEYIYQRRPEYRHLASGDAVAVQKLSLNSCKSFKWFMTKIAWDLPKFYPPVEPP 456
Qy 444 -GFGMLONKGLTDYCFDYNPDENQIVGHQVILYLC-HGMQO-----NQFEYTSQKEI 496
Db 457 AAANGEIRNVG-TGLCAD---TKHGALGSLRLEGCVGRGEAANNMQVFTFTWREDI 511
Qy 497 RYN--THQEGCIAVEAGMDTLIMHLCETAPENQKFIQEDGSLFHEQSKKCVQAAARKE 554
Db 512 RPDGPQHTKCFDAISHTSPVTLYDC-HSMKGNQLWKYRKDKTLYHPVSGSCM-----564
Qy 555 SSSDFVPLLRDCTNSDH-----QKWFKE 577
Db 565 -----DCSESDHRIFMNTCNPSLTQOWLFE 590
```

## RESULT 13

```
US-09-795-926-31
; Sequence 31, Application US/09795926
; Patent No. 6555669
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Hilbun, Erin
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; APPLICANT: Walke, D. Wade
; APPLICANT: Wilganowski, Nathaniel L.
; APPLICANT: Hu, Yi
; APPLICANT: Kiege, James Alvin
; APPLICANT: Potter, David George
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
; FILE REFERENCE: POLYNUCLEOTIDES ENCODING THE SAME
; CURRENT APPLICATION NUMBER: US/09/795,926
; FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/185,920
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 60/186,558
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/191,849
```

```
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 506
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-795-926-31
```

## Query Match 31.2%; Score 974.5; DB 4; Length 506;

```
Best Local Similarity 40.5%; Pred. No. 2.6e-91;
Matches 207; Conservative 84; Mismatches 167; Indels 53; Gaps 14;
```

```
Qy 95 EESVRLHQINIVLSDRISLHRRLPERWNPCKEKKYDYNLPRTSVILAFYNEAWSTLLR 154
Db 8 DQAYRENGFNIVYSDKISLNSRLPDIRHPNCNKRY-LETLPNTSIIPFHNEGWSSLLR 66
Qy 155 TVSVLETSPIILLEVILVDDYSDEHLEKRLANLGLPKVRLIRANKREGVLRL 214
Db 67 TVHSLNRPPELVAIVLVDPSDEHLEKRLANLGLPKVRLIRANKREGVLRL 126
Qy 215 GASAARGDVLTFDCHCEHGLEPQLQRIHEESAVVCPVIDIDMNTTFYLGNSGP 274
Db 127 GASVATGDTITFLDSHCEANVNWLPPLLDRIARNRKTIVCPMIDVIDHDDPRYETQAGDA 186
Qy 275 QIGGFWRLVFTWHTVPERIRIMQSPVDVIRSPMTAGGLFAVSKKYFYLGSDYDTGMEV 334
Db 187 MRGAPDWEY--KRIPPELQKADPSDPFSPVMAGGLFAVDRKWFELGGYDPLGLEI 244
Qy 335 WGGENLEFSPRIWOCGGVLETHPCSHVGHVFPKQAPYS--RNKALA-NSVRAAEVWMD 391
Db 245 WGGQYEISFKVMWCGGRMEDIPCSRVGHIYRYKVPYKVPAGVSLARNLKRVAEVMWDEY 304
Qy 392 KELYHNRPRARLEPFGDVTERRKQLDKLQCKDFKWFLETV---YPELHVPEDRP-CFFG 447
Db 305 AEYIYQRRPEYRHLASGDAVAVQKLSLNSCKSFKWFMTKIAWDLPKFYPPVEPPAAAG 364
Qy 448 MLQNKGLTDYCFDYNPDENQIVGHQVILYLC-HGMQO-----NQFEYTSQKEIRYN-- 499
Db 365 EIRNVG-TGLCAD---TKHGALGSLRLEGCVGRGEAANNMQVFTFTWREDIRPGDP 419
Qy 500 THQEGCIAVEAGMDTLIMHLCETAPENQKFIQEDGSLFHEQSKKCVQAAARKESSDSF 559
Db 420 QHTKKCFDAISHTSPVTLYDC-HSMKGNQLWKYRKDKTLYHPVSGSCM-----467
Qy 560 VPLLRDCTNSDH-----QKWFKE 577
Db 468 -----DCSESDHRIFMNTCNPSLTQOWLFE 493
```

## RESULT 14

```
US-09-795-926-41
; Sequence 41, Application US/09795926
; Patent No. 6555669
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Hilbun, Erin
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; APPLICANT: Walke, D. Wade
; APPLICANT: Wilganowski, Nathaniel L.
; APPLICANT: Hu, Yi
; APPLICANT: Kiege, James Alvin
; APPLICANT: Potter, David George
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
; FILE REFERENCE: POLYNUCLEOTIDES ENCODING THE SAME
; CURRENT APPLICATION NUMBER: US/09/795,926
; FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/185,920
```

```
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 60/186,558
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/191,849
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41
; LENGTH: 631
; TYPE: PRT
; ORGANISM: homo sapiens
; US-09-795-926-41

Query Match      31.0%; Score 967; DB 4; Length 631;
Best Local Similarity 34.3%; Pred. No. 2.2e-90;
Matches 231; Conservative 95; Mismatches 197; Indels 142; Gaps 20;

QY      8 RRCPRELRGREALVLLALLALAGLGLSVLRAQR-----GAGAGAAEPGPPRTPRGRR 61
DB      2 RRKEKRLQVALVLAALVLLPNVGLMALYRERQDGTGGGAAVA----- 48
QY      62 EPVMPRPVPANALGARG-----EAVRLQLQ-----GEELR----- 92
DB      49 -----PAAGGSHSRQKTFFLGDGQKLDKDHKEAIRDAQRVGNAGEQRPYPMT 99
QY      93 ---LQESVRLHQINYLSDRISLHRRLPERNWPLCKEKKYDYNLPTSVIIAFYNEAW 149
DB      100 DAERVQDAYRENGFNIVYSDKISLNSRLPDTRHPNCNKRY-LETLPNTSIIIPFHNEGW 158
QY      150 STLLRTVYSLVETSPDLLLEEVILVDDYSDREHLKERLANSLSGLPKVRLIRANKRGIV 209
DB      159 SLLRTVHSLNRPPELVAEIVLDDFSDREHLKPKLEDYMAFPFVSRILRTKKREGLI 218
QY      210 RARLLGASAAAGDVLTFDCHCECHGWLPELQLRIHEESAVVCPVIDVIDMNTFYLG 269
DB      219 RTRMLGASVATGDIITLDSHCEANVNLPLLDRIARNRKTIVCPMIDVIDHDDFRYET 278
QY      270 NSGEPOIGGDFWRILVFTWHTVPERERIRMSQSPDVIRSPMTAGGLFAVSKYFEYLGSDY 329
DB      279 QAGDAMRGAFDWEMY--KRIPPELQKADPSDPFESVWAGGLFAVDRKWFELGGYD 336
QY      330 TGMVWNGENLEFSFR-----IWQCGVLETHPCSH 360
DB      337 PGLIEWGEQVEISFKGLHMLPRLVSNWPQAVFLPRAPNMLAQVWMCGRMEDIPCSR 396
QY      361 VGHVFPKQAPYS--RNKALA-NSVRAAEVWMDPEFKELYHNRPRARLEPFGDVTTERKQLR 417
DB      397 VGHYRYKYVPYKYPAGVSLARNLKRVAEVMWDEYAEIYQRRPEYRHLISAGDVAVQKKLR 456
QY      418 DKLQCKDFKWFLETV---YPELHVPEDRP-GFFGMLQNKGLTDYCFDYNPPDENQIVGHQ 473
DB      457 SSLNCKSFKWFMTKIAMDLPKFYPPEVPPAAANGAIRNVG-TGLCAD---TKHGALGSP 511
QY      474 VILYLC-HGMGQ-----NQFEYTSQKEIRYN--THOPEGCIAVEAGMDTLIMHLCETAPENQK 525
DB      512 LRLEGCVRGGEAAANNQVFTFTWRDIRPGDPQHTKKCFDAISHTSPVTLYDC-HSM 570
QY      526 PENQKFIQBDGSLFHEQSKKCVQAARKESSDSFVPLLRDCTNSDH-----Q 572
DB      571 KGNQLWKYRKDKTLYHPVSGSCM-----DCSESDDRIFMNTCNPSLTLQ 614
QY      573 KWFPK 577
DB      615 QWLFE 619
```

## RESULT 15

```
US-09-795-926-29
; Sequence 29, Application US/09795926
; Patent No. 6555669
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Hilburn, Brin
```

```
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; APPLICANT: Walke, D. Wade
; APPLICANT: Wilganowski, Nathaniel L.
; APPLICANT: Hu, Yi
; APPLICANT: Kieke, James Alvin
; APPLICANT: Potter, David George
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
; FILE REFERENCE: LEX-0144-USA
; CURRENT APPLICATION NUMBER: US/09/795,926
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/185,920
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 60/186,558
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/191,849
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 535
; TYPE: PRT
; ORGANISM: homo sapiens
; US-09-795-926-29
```

Query Match 30.4%; Score 950; DB 4; Length 535;

Best Local Similarity 38.3%; Pred. No. 9.4e-89;

Matches 207; Conservative 84; Mismatches 167; Indels 82; Gaps 15;

```
QY      95 EESVRLHQINYLSDRISLHRRLPERNWPLCKEKKYDYNLPTSVIIAFYNEAWSTLLR 154
DB      8 DQAYRENGFNIVYSDKISLNSRLPDTRHPNCNKRY-LETLPNTSIIIPFHNEGWSELLR 66
QY      155 TVYSVLETSPDLLLEEVILVDDYSDREHLKERLANSLSGLPKVRLIRANKRGIVRARLL 214
DB      67 TVHSLNRPPELVAEIVLDDFSDREHLKPKLEDYMAFPFVSRILRTKKREGIIRTML 126
QY      215 GASAAAGDVLTFDCHCECHGWLPELQLRIHEESAVVCPVIDVIDMNTFYLGNSGEP 274
DB      127 GASVATGDIITLDSHCEANVNLPLLDRIARNRKTIVCPMIDVIDHDDFRYETQAGDA 186
QY      275 QIGGDFWRILVFTWHTVPERERIRMSQSPDVIRSPMTAGGLFAVSKYFEYLGSDYTGMEV 334
DB      187 MRGAFDWEY--KRIPPELQKADPSDPFESVWAGGLFAVDRKWFELGGYDPLGLEI 244
QY      335 WGGENLEFSFR-----IWQCGVLETHPCSHVGHVF 365
DB      245 WGGQVEISFKGLHMLPRLVSNWPQAVFLPRAPNMLAQVWMCGRMEDIPCSRUGHY 304
QY      366 PKQAPYS--RNKALA-NSVRAAEVWMDPEFKELYHNRPRARLEPFGDVTTERKQLRKLQC 422
DB      305 RKVVPYKYPAGVSLARNLKRVAEVMWDEYAEIYQRRPEYRHLISAGDVAVQKKLSLNC 364
QY      423 KDFKWFLETV---YPELHVPEDRP-GFFGMLQNKGLTDYCFDYNPPDENQIVGHQVILYL 478
DB      365 KSKFWMTKIAMDLPKFYPPEVPPAAANGAIRNVG-TGLCAD---TKHGALGSPRLLEG 419
QY      479 C-HGMGQ-----NQFEYTSQKEIRYN--THOPEGCIAVEAGMDTLIMHLCETAPENQK 530
DB      420 CVRGGEAAANNQVFTFTWRDIRPGDPQHTKKCFDAISHTSPVTLYDC-HSMKGNQL 478
QY      531 FILQEDGSLFHEQSKKCVQAARKESSDSFVPLLRDCTNSDH-----QKWF 577
DB      479 WKYRKDKTLYHPVSGSCM-----DCSESDDRIFMNTCNPSLTLQWLFE 522
```

Search completed: December 4, 2003, 19:44:49

Job time : 23 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: December 4, 2003, 19:43:48 ; Search time 36 Seconds  
(without alignments)  
3001.570 Million cell updates/sec

Title: US-10-074-527-2

Perfect score: 3124

Sequence: 1 MWGTTARRCPRELRRGREA.....LLRDCTNSDHQKWFKERM 581

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 684280 seqs, 185983659 residues

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

```
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pap:*
2: /cgn2_6/ptodata/2/pubpaa/PTCT_NEW_PUB.pap:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pap:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pap:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pap:*
6: /cgn2_6/ptodata/2/pubpaa/PTCTUS_PUBCOMB.pap:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pap:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pap:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pap:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pap:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pap:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pap:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pap:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pap:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pap:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pap:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pap:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pap:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3124	100.0	581	14	US-10-074-527-2
2	3076	98.5	581	12	Sequence 2, Appli
3	1675	53.6	578	14	Sequence 58, Appli
4	1369	43.8	256	9	US-10-074-527-8
5	1369	43.8	256	11	US-09-925-299-992
6	1309	41.9	316	15	US-09-925-299-992
7	1304	41.7	289	11	US-10-106-698-4875
8	1176	37.6	626	14	Sequence 4875, Ap
9	1142.5	36.6	633	15	Sequence 95, Appli
10	1122.5	35.9	556	12	US-10-001-851-27
11	1109	35.5	559	14	US-10-205-823-146
12	1101.5	35.3	559	14	Sequence 146, App
13	1101.5	35.3	559	14	Sequence 62, Appli
14	1100.5	35.2	559	9	US-10-001-851-20
15	1097	35.1	559	12	Sequence 20, Appli
					Sequence 1006, Ap
					Sequence 23, Appli
					Sequence 36, Appli

## ALIGNMENTS

## RESULT 1

US-10-074-527-2

; Sequence 2, Application US/10074527

; Publication No. US20020142426A1

; GENERAL INFORMATION:

; APPLICANT: Olandt, Peter J.

; APPLICANT: Meyers, Rachel E.

; APPLICANT: Galvin, Katherine A.

; APPLICANT: Millennium Pharmaceuticals Inc.

; TITLE OF INVENTION: 33945, A Human Glycosyltransferase and

; TITLE OF INVENTION: Uses Therefor

; FILE REFERENCE: MPI2001-0181RCP1(W)

; CURRENT APPLICATION NUMBER: US/10/074,527

; CURRENT FILING DATE: 2002-02-12

; PRIOR APPLICATION NUMBER: 60/269202

; PRIOR FILING DATE: 2001-02-15

; NUMBER OF SEQ ID NOS: 9

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 2

; LENGTH: 581

; TYPE: PRT

; ORGANISM: homo sapiens

; US-10-074-527-2

Query Match 100.0%; Score 3124; DB 14; Length 581;

Best Local Similarity 100.0%; Pred. No. 7.8e-283;

Matches 581; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MWGTTARRCPRELRRGREA.....LLRDCTNSDHQKWFKERM 581

Db 1 MWGTTARRCPRELRRGREA.....LLRDCTNSDHQKWFKERM 581

Qy 61 REPVMPRPVPANALGARGAVRLQGEELRQESVRLHQINIVYLSRISLHRLRPER 120

Db 61 REPVMPRPVPANALGARGAVRLQGEELRQESVRLHQINIVYLSRISLHRLRPER 120

Qy 121 WNPICKEKKYDYNLPRTSVIIAFYNEAWSTLTRTVYVLETSPDILLEVLVDYSDR 180

Db 121 WNPICKEKKYDYNLPRTSVIIAFYNEAWSTLTRTVYVLETSPDILLEVLVDYSDR 180

Qy 121 WNPICKEKKYDYNLPRTSVIIAFYNEAWSTLTRTVYVLETSPDILLEVLVDYSDR 180

Db 121 WNPICKEKKYDYNLPRTSVIIAFYNEAWSTLTRTVYVLETSPDILLEVLVDYSDR 180

Qy 121 WNPICKEKKYDYNLPRTSVIIAFYNEAWSTLTRTVYVLETSPDILLEVLVDYSDR 180

Db 121 WNPICKEKKYDYNLPRTSVIIAFYNEAWSTLTRTVYVLETSPDILLEVLVDYSDR 180

Db 121 WNPLCKEKKYDYNLPTSVIIAFYNEAWSTLLRTVSVLETSPIILLEEVILVDDYSR 180  
Qy 181 EHLKERLANELSGLPKVRILIRANKREGVRLVRLGASAGDVLTFDCHCEHEGWLEP 240  
Db 181 EHLKERLANELSGLPKVRILIRANKREGVRLVRLGASAGDVLTFDCHCEHEGWLEP 240  
Qy 241 LLQRIHEESAVVCPVIDVDWNTFEYLGNSGEPQIGGFOWRLVFTWHTVPERERIRMOS 300  
Db 241 LLQRIHEESAVVCPVIDVDWNTFEYLGNSGEPQIGGFOWRLVFTWHTVPERERIRMOS 300  
Qy 301 PVDVIRPTMAGGLFAVSKYFEYLGSDYDTGMEVWNGENLEFSPRIWQCGVLETHPCSH 360  
Db 301 PVDVIRPTMAGGLFAVSKYFEYLGSDYDTGMEVWNGENLEFSPRIWQCGVLETHPCSH 360  
Qy 361 VGHVFPKQAPYSRNKALANSVRAAEVWMDPEFKELYHNRPRARLEPFGDVTERRKQDRDL 420  
Db 361 VGHVFPKQAPYSRNKALANSVRAAEVWMDPEFKELYHNRPRARLEPFGDVTERRKQDRDL 420  
Qy 421 QCKDFKWFLETVPELVHPEDRPGFGMLQNKGLTDYCFDYNPPDENQIVGHQVILYLCH 480  
Db 421 QCKDFKWFLETVPELVHPEDRPGFGMLQNKGLTDYCFDYNPPDENQIVGHQVILYLCH 480  
Qy 481 GMGQNFPEYTSQKEIRYNTHQPEGCIATVAGMDTLIMHLCETAPENQKFIQEDGSLF 540  
Db 481 GMGQNFPEYTSQKEIRYNTHQPEGCIATVAGMDTLIMHLCETAPENQKFIQEDGSLF 540  
Qy 541 HEQSKKCVQAARKESDSFVPLLRDCTNSDHQKWFKERM 581  
Db 541 HEQSKKCVQAARKESDSFVPLLRDCTNSDHQKWFKERM 581

## RESULT 2

US-10-292-896-58

; Sequence 58, Application US/10292896

; Publication No. US20030186850A1

; GENERAL INFORMATION:

; APPLICANT: HASSAN, Helle

; APPLICANT: REIS, Celso A.

; APPLICANT: BENNETT, Eric P.

; APPLICANT: CLAUSEN, Henrik

; TITLE OF INVENTION: METHODS TO IDENTIFY AGENTS MODULATING FUNCTIONS OF POLYPEPTIDE GA

; TITLE OF INVENTION: TRANSFERASES, PHARMACEUTICAL COMPOSITIONS COMPRISING SUCH AGENTS

; TITLE OF INVENTION: SUCH AGENTS FOR PREPARING MEDICAMENTS

; FILE REFERENCE: 4305/1H154-US3

; CURRENT APPLICATION NUMBER: US/10/292,896

; CURRENT FILING DATE: 2003-03-17

; PRIOR APPLICATION NUMBER: US 60/425,204

; PRIOR FILING DATE: 2002-11-08

; PRIOR APPLICATION NUMBER: PCT/DK01/00328

; PRIOR FILING DATE: 2001-05-10

; PRIOR APPLICATION NUMBER: US 60/203,331

; PRIOR FILING DATE: 2000-05-11

; NUMBER OF SEQ ID NOS: 130

; SOFTWARE: Patent in version 3.1

; SEQ ID NO 58

; LENGTH: 581

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-292-896-58

Query Match 98.5%; Score 3076; DB 12; Length 581;

Best Local Similarity 98.5%; Pred. No. 2.4e-278;

Matches 572; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 1 MWGTARRRCPRELRRGREALLVLLALAGLSVLRQAORAGAGAAEPGPPTPRGR 60  
Db 1 MWGTARRRCPRELRRGREALLVLLALAGLSVLRQAORAGAGAAEPGPPTPRGR 60

Qy 61 REPWMPRPVPPANALGARGEAVRLQOGEEURLQOESVRLHQINILSDRISLHRLPER 120  
Db 61 REPWMPRPVPPANALGARGEAVRLQOGEEURLQOESVRLHQINILSDRISLHRLPER 120

Qy 121 WNPLCKEKKYDYNLPTSVIIAFYNEAWSTLLRTVSVLETSPIILLEEVILVDDYSR 180

Db 121 WNPLCKEKKYDYNLPTSVIIAFYNEAWSTLLRTVSVLETSPIILLEEVILVDDYSR 180  
Qy 181 EHLKERLANELSGLPKVRILIRANKREGVRLVRLGASAGDVLTFDCHCEHEGWLEP 240  
Db 181 EHLKERLANELSGLPKVRILIRANKREGVRLVRLGASAGDVLTFDCHCEHEGWLEP 240  
Qy 241 LLQRIHEESAVVCPVIDVDWNTFEYLGNSGEPQIGGFOWRLVFTWHTVPERERIRMOS 300  
Db 241 LLQRIHEESAVVCPVIDVDWNTFEYLGNSGEPQIGGFOWRLVFTWHTVPERERIRMOS 300  
Qy 301 PVDVIRPTMAGGLFAVSKYFEYLGSDYDTGMEVWNGENLEFSPRIWQCGVLETHPCSH 360  
Db 301 PVDVIRPTMAGGLFAVSKYFEYLGSDYDTGMEVWNGENLEFSPRIWQCGVLETHPCSH 360  
Qy 361 VGHVFPKQAPYSRNKALANSVRAAEVWMDPEFKELYHNRPRARLEPFGDVTERRKQDRDL 420  
Db 361 VGHVFPKQAPYSRNKALANSVRAAEVWMDPEFKELYHNRPRARLEPFGDVTERRKQDRDL 420  
Qy 421 QCKDFKWFLETVPELVHPEDRPGFGMLQNKGLTDYCFDYNPPDENQIVGHQVILYLCH 480  
Db 421 QCKDFKWFLETVPELVHPEDRPGFGMLQNKGLTDYCFDYNPPDENQIVGHQVILYLCH 480  
Qy 481 GMGQNFPEYTSQKEIRYNTHQPEGCIATVAGMDTLIMHLCETAPENQKFIQEDGSLF 540  
Db 481 GMGQNFPEYTSQKEIRYNTHQPEGCIATVAGMDTLIMHLCETAPENQKFIQEDGSLF 540  
Qy 541 HEQSKKCVQAARKESDSFVPLLRDCTNSDHQKWFKERM 581  
Db 541 HEQSKKCVQAARKESDSFVPLLRDCTNSDHQKWFKERM 581

## RESULT 3

US-10-074-527-8

; Sequence 8, Application US/10074527

; Publication No. US20020142426A1

; GENERAL INFORMATION:

; APPLICANT: Olandt, Peter J.

; APPLICANT: Meyers, Rachel E.

; APPLICANT: Gaivin, Katherine A.

; APPLICANT: Millennium Pharmaceuticals Inc.

; TITLE OF INVENTION: 33945, A Human Glycosyltransferase and

; TITLE OF INVENTION: Uses Therefor

; FILE REFERENCE: MP12001-018P1RCP1(M)

; CURRENT APPLICATION NUMBER: US/10/074,527

; CURRENT FILING DATE: 2002-02-12

; PRIOR APPLICATION NUMBER: 60/269202

; PRIOR FILING DATE: 2001-02-15

; NUMBER OF SEQ ID NOS: 9

; SOFTWARE: Fast-Seq for Windows Version 4.0

; SEQ ID NO 8

; LENGTH: 578

; TYPE: PRT

; ORGANISM: Mus musculus

US-10-074-527-8

Query Match 53.6%; Score 1675; DB 14; Length 578;

Best Local Similarity 57.9%; Pred. No. 2e-147;

Matches 330; Conservative 78; Mismatches 144; Indels 18; Gaps 8;

Qy 23 VLLALLALAGL-----GSVLRQAORAGAGAAEPGPPTPRGR-----PVMPRPVPA 72  
Db 13 LLLALLTLAVLVEFVSSTLYASFGAG-GARELGPRLPDLDTREEDLSQPLYIKPADS 71

Qy 73 NALGARGEAVRLQOGEEURLQOESVRLHQINILSDRISLHRLPERWNPCKEKKYD 132  
Db 72 HALGEWGRASKLQNEGELKQOSELIERVAINILSDRISLHRIEDKMYECKAKKPHY 131

Qy 133 DNLPTSVIIAFYNEAWSTLLRTVSVLETSPIILLEEVILVDDYSREHLKERLANELS 192  
Db 132 RSLPTTSVIIAFYNEAWSTLLRTVSVLETSPIILLEEVILVDDYSREHLKERLANELS 191

Qy 193 GLPKVRLIRANKREGVRLVRLGASAGDVLTFDCHCEHEGWLEPPLQRIHEESAV 252



Db 192 NLERVLRIRNKREGLVARLIGATFATGDLVLFDCHECNCNTGWLPLELIERISRDETAI 251  
Qy 253 VCPVIDVDWNTFEYLGNSGEPQIGGDFWRLVFTWHTVPERIRMQSPVDVIRSPTMAG 312  
Db 252 VCPVIDIDWNTFEFYMQTGEMPGDFWRLTFQWHSVPKHERDRRTSRIDPIRSPTMAG 311  
Qy 313 GLFAVSKKYFEYLGSDYDTGMEVWGENLEFSPRIWQCGGVLETHPCSHVGHVFPKQAPYS 372  
Db 312 GLFAVSKKYFOYLGTYDTGMEVWGENLELSFRWQCGGKLEIHPCHSHVGHVFPKRAPYA 371  
Qy 373 RNKALANSVRAAEVWMDFEKELYHYRNPRLARLEPFGDVTERRKOLRDKLOCKDFKWFLETV 432  
Db 372 RPNFLQNTARAAEVWMDYKEHFYRNPRLARLEPFGDVTERRKOLRDKCKSDFWYLVNV 431  
Qy 433 YPELVHPEDRPGFGLMKNGLTDYCFDYNPPDENQIVGHQVILYLCHGMGQNOFFEYTS 492  
Db 432 FSNLHVPEDRPGFGLMKNGLTDYCFDYNPPDENQIVGHQVILYLCHGMGQNOFFEYTS 490  
Qy 493 QKEIRYNTHOPEGCIAGVAGMDTLIMHLCSE---TAPENQKFILOBDGSLFHSQSKCVQ 549  
Db 491 NKEIRFNS-VTELCAEYPOQKYVGMQNCPKDGLFVPVNIWHFKEDGTIFHPHTRLCLS 549  
Qy 550 AAKESSDSFVPLLRDCTNSD-HQKWFKE 578  
Db 550 AYRTAEGRPVH-MKTCDAIDKQNLWRFEK 578

## RESULT 4

US-09-925-299-992  
; Sequence 992, Application US/09925299  
; Patent No. US20020055627A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA102  
; CURRENT APPLICATION NUMBER: US/09/925,299  
; PRIOR FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05883  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 1556  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 992  
; LENGTH: 256  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (229)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-299-992  
Query Match 43.8%; Score 1369; DB 9; Length 256;  
Best Local Similarity 99.6%; Pred. No. 2.3e-119;  
Matches 247; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 238 LEPLQRIHEESAVVCPVIDIDWNTFEYLGNSGEPQIGGDFWRLVFTWHTVPERIR 297  
Db 6 LEPLQRIHEESAVVCPVIDIDWNTFEYLGNSGEPQIGGDFWRLVFTWHTVPERIR 65  
Qy 298 MQSPVDVIRSPTMAGGLFAVSKKYFEYLGSDYDTGMEVWGENLEFSFRIWQCGGVLETHP 357  
Db 66 MQSPVDVIRSPTMAGGLFAVSKKYFEYLGSDYDTGMEVWGENLEFSFRIWQCGGVLETHP 125  
Qy 358 CSHVGHVFPKQAPYSRNKALANSVRAAEVWMDFEKELYHYRNPRLARLEPFGDVTERRKOLR 417  
Db 126 CSHVGHVFPKQAPYSRNKALANSVRAAEVWMDFEKELYHYRNPRLARLEPFGDVTERRKOLR 185  
Qy 418 DKLOCKDFKWFLETVPELVHPEDRPGFGLMKNGLTDYCFDYNPPDENQIVGHQVILY 477  
Db 186 DKLOCKDFKWFLETVPELVHPEDRPGFGLMKNGLTDYCFDYNPPDENQIVGHQVILY 245

US-09-925-299-992  
Query Match 43.8%; Score 1369; DB 9; Length 256;  
Best Local Similarity 99.6%; Pred. No. 2.3e-119;  
Matches 247; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 238 LEPLQRIHEESAVVCPVIDIDWNTFEYLGNSGEPQIGGDFWRLVFTWHTVPERIR 297  
Db 6 LEPLQRIHEESAVVCPVIDIDWNTFEYLGNSGEPQIGGDFWRLVFTWHTVPERIR 65  
Qy 298 MQSPVDVIRSPTMAGGLFAVSKKYFEYLGSDYDTGMEVWGENLEFSFRIWQCGGVLETHP 357  
Db 66 MQSPVDVIRSPTMAGGLFAVSKKYFEYLGSDYDTGMEVWGENLEFSFRIWQCGGVLETHP 125  
Qy 358 CSHVGHVFPKQAPYSRNKALANSVRAAEVWMDFEKELYHYRNPRLARLEPFGDVTERRKOLR 417  
Db 126 CSHVGHVFPKQAPYSRNKALANSVRAAEVWMDFEKELYHYRNPRLARLEPFGDVTERRKOLR 185  
Qy 418 DKLOCKDFKWFLETVPELVHPEDRPGFGLMKNGLTDYCFDYNPPDENQIVGHQVILY 477  
Db 186 DKLOCKDFKWFLETVPELVHPEDRPGFGLMKNGLTDYCFDYNPPDENQIVGHQVILY 245

## RESULT 6

US-10-106-698-4875  
; Sequence 4875, Application US/10106698  
; Publication No. US20030109690A1  
; GENERAL INFORMATION:  
; APPLICANT: Ruben et al.  
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides  
; FILE REFERENCE: PA005P1  
; CURRENT APPLICATION NUMBER: US/10/106,698  
; CURRENT FILING DATE: 2002-03-27  
; PRIOR APPLICATION NUMBER: PCT/US00/26524  
; PRIOR FILING DATE: 2000-09-28  
; PRIOR APPLICATION NUMBER: US 60/157,137  
; PRIOR FILING DATE: 1999-09-29  
; PRIOR APPLICATION NUMBER: US 60/163,280  
; PRIOR FILING DATE: 1999-11-03

Qy 478 LCHGMGQN 485  
Db 246 LCHGMGQN 253

## RESULT 5

US-09-925-299-992  
; Sequence 992, Application US/09925299  
; Publication No. US20030040617A9  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA102  
; CURRENT APPLICATION NUMBER: US/09/925,299  
; PRIOR FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05883  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 1556  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 992  
; LENGTH: 256  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (229)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-299-992  
Query Match 43.8%; Score 1369; DB 11; Length 256;  
Best Local Similarity 99.6%; Pred. No. 2.3e-119;  
Matches 247; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 238 LEPLQRIHEESAVVCPVIDIDWNTFEYLGNSGEPQIGGDFWRLVFTWHTVPERIR 297  
Db 6 LEPLQRIHEESAVVCPVIDIDWNTFEYLGNSGEPQIGGDFWRLVFTWHTVPERIR 65  
Qy 298 MQSPVDVIRSPTMAGGLFAVSKKYFEYLGSDYDTGMEVWGENLEFSFRIWQCGGVLETHP 357  
Db 66 MQSPVDVIRSPTMAGGLFAVSKKYFEYLGSDYDTGMEVWGENLEFSFRIWQCGGVLETHP 125  
Qy 358 CSHVGHVFPKQAPYSRNKALANSVRAAEVWMDFEKELYHYRNPRLARLEPFGDVTERRKOLR 417  
Db 126 CSHVGHVFPKQAPYSRNKALANSVRAAEVWMDFEKELYHYRNPRLARLEPFGDVTERRKOLR 185  
Qy 418 DKLOCKDFKWFLETVPELVHPEDRPGFGLMKNGLTDYCFDYNPPDENQIVGHQVILY 477  
Db 186 DKLOCKDFKWFLETVPELVHPEDRPGFGLMKNGLTDYCFDYNPPDENQIVGHQVILY 245

US-09-925-299-992  
Query Match 43.8%; Score 1369; DB 9; Length 256;  
Best Local Similarity 99.6%; Pred. No. 2.3e-119;  
Matches 247; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 238 LEPLQRIHEESAVVCPVIDIDWNTFEYLGNSGEPQIGGDFWRLVFTWHTVPERIR 297  
Db 6 LEPLQRIHEESAVVCPVIDIDWNTFEYLGNSGEPQIGGDFWRLVFTWHTVPERIR 65  
Qy 298 MQSPVDVIRSPTMAGGLFAVSKKYFEYLGSDYDTGMEVWGENLEFSFRIWQCGGVLETHP 357  
Db 66 MQSPVDVIRSPTMAGGLFAVSKKYFEYLGSDYDTGMEVWGENLEFSFRIWQCGGVLETHP 125  
Qy 358 CSHVGHVFPKQAPYSRNKALANSVRAAEVWMDFEKELYHYRNPRLARLEPFGDVTERRKOLR 417  
Db 126 CSHVGHVFPKQAPYSRNKALANSVRAAEVWMDFEKELYHYRNPRLARLEPFGDVTERRKOLR 185  
Qy 418 DKLOCKDFKWFLETVPELVHPEDRPGFGLMKNGLTDYCFDYNPPDENQIVGHQVILY 477  
Db 186 DKLOCKDFKWFLETVPELVHPEDRPGFGLMKNGLTDYCFDYNPPDENQIVGHQVILY 245



## RESULT 9

US-10-205-823-146  
; Sequence 146, Application US/10205823  
; Publication No. US20030108963A1  
; GENERAL INFORMATION:  
; APPLICANT: Schlegel, Robert  
; APPLICANT: Monahan, John E.  
; APPLICANT: Endege, Wilson O.  
; APPLICANT: Gannavarapu, Manjula  
; APPLICANT: Gorbacheva, Bella  
; APPLICANT: Hoersch, Sebastian  
; APPLICANT: Kamatkar, Shubhangi  
; APPLICANT: Wonshey, Angela M.  
; APPLICANT: Glatt, Karen  
; APPLICANT: Zhao, Xumei  
; APPLICANT: Anderson, Dustin  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND  
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
; TITLE OF INVENTION: THERAPY OF PROSTATE CANCER  
; FILE REFERENCE: MRI-044  
; CURRENT APPLICATION NUMBER: US/10/205,823  
; CURRENT FILING DATE: 2002-07-25  
; PRIOR FILING DATE: 60/307,982  
; PRIOR FILING DATE: 2001-07-25  
; PRIOR APPLICATION NUMBER: 60/314,356  
; PRIOR FILING DATE: 2001-08-22  
; PRIOR APPLICATION NUMBER: 60/325,020  
; PRIOR FILING DATE: 2001-09-25  
; PRIOR APPLICATION NUMBER: 60/341,746  
; PRIOR FILING DATE: 2001-12-12  
; PRIOR APPLICATION NUMBER: 60/362,158  
; PRIOR FILING DATE: 2002-03-05  
; NUMBER OF SEQ ID NOS: 455  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 146  
; LENGTH: 633  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-205-823-146

Query Match 36.6%; Score 1142.5; DB 15; Length 633;  
Best Local Similarity 44.9%; Pred. No. 1.3e-97;  
Matches 244; Conservative 90; Mismatches 163; Indels 47; Gaps 14;  
Qy 62 EPVMPRPVPANALGARGEAVR-LQAGEELRLQESVRLHQLINIVLSDRISLHRL-PE 119  
Db 108 KPVLDLPQDSNAGSGAKFTTNLSVEEQEKERGEAKCFNAPASDRISLHRLDLPD 167  
Qy 120 RWNPLCKEKYD-YDNLPRTSVIAFYNEAWSTLLRTVYSVLETSPIILLBEVILVDDYS 178  
Db 168 TRPECEIQKPKRCPPLPTTSVIVFHNEAWSTLLRTVHSVLYSSPAILLKEIILVDDAS 227  
Qy 179 DREHLKERLANELSGLPKRLIRANKREGLVRALLGASARGVDVLPDCHCECHGWL 238  
Db 228 VDEYLDKLEVYKQFSIVKIVRQERKGLTARLLGATATATLTFLDAHCECFYGL 287  
Qy 239 EPLLQRIHEESAVVCPVIDVDMNTFEYLGNSGEP-----QIGCFDWRVLTWHTVPER 293  
Db 288 EPLLARLAENYAVVSFDIASIDNTPEF--NKSPVGSNNRNPNFWSLSFGWESLPDH 345  
Qy 294 ERIRMQSPVDVIRSPPTMAGGLFVSKKYFEYGLSYDTGMVWGGENLEFSGFRWQCQGV 353  
Db 346 EKQRKDETPYIKPTPTAGGLFSSKEYFEYIGSYDEMEIWGGENIEMSFVWQCQGL 405  
Qy 354 ETHPCSHVGHVPPKQAPYSRKA-----LANSVRAAEVWMDPEFKELYHNRNPA-----RLE 405  
Db 406 EIMPCSVGVHVFRRSKSPHSPFKGTQVIARNQVRLAEVWMDPEFKELYHNRNPA-----RLE 465  
Qy 406 PFGDVTERRKQLRDLCKQKFWLETYVPELHVPEDEPFGFGLMQLONKGLTDYCFDYNPPD 465  
Db 466 AFGSLKRRFEIKHLRCKNFTWYLNNTYPEVYVPELHVPEDEPFGFGLMQLONKGLTDYCFDYNPPD 521

Qy 466 ENQIVGHQVILYLCHGMGQNFQFYTSQKEIRYN-----THQPEGCIAVEA-----GMDT 515  
Db 522 ENNGGKPLIMYTCHGLGNGQFYEYSAQHEIRHNIQKELCLHAAQGLVQLKACTYKGHKT 581  
Qy 516 LIMHLCETAPENQKFTLQEDGSLFHEQSKKCVQARKESSDSPVPLLRDCTNSDH-QKW 574  
Db 582 VW-----TGEQIWEIQDKQLLYNPFLLKWSANGEH-----PSLVSNPSPDLPQKW 627  
Qy 575 FPKK 578  
Db 628 ILSQ 631  
RESULT 10  
US-10-292-896-62  
; Sequence 62, Application US/10292896  
; Publication No. US20030186850A1  
; GENERAL INFORMATION:  
; APPLICANT: HASSAN, Helle  
; APPLICANT: REIS, Celso A.  
; APPLICANT: BENNETT, Eric P.  
; APPLICANT: CLAUSEN, Henrik  
; TITLE OF INVENTION: METHODS TO IDENTIFY AGENTS MODULATING FUNCTIONS OF POLYPEPTIDE GAI  
; TITLE OF INVENTION: TRANSFERASES, PHARMACEUTICAL COMPOSITIONS COMPRISING SUCH AGENTS  
; TITLE OF INVENTION: SUCH AGENTS FOR PREPARING MEDICAMENTS  
; FILE REFERENCE: 4305/1H154-US3  
; CURRENT APPLICATION NUMBER: US/10/292,896  
; CURRENT FILING DATE: 2003-03-17  
; PRIOR FILING DATE: 60/425,204  
; PRIOR APPLICATION NUMBER: PCT/DK01/00328  
; PRIOR FILING DATE: 2002-11-08  
; PRIOR FILING DATE: 2001-05-10  
; PRIOR APPLICATION NUMBER: US 60/203,331  
; PRIOR FILING DATE: 2000-05-11  
; NUMBER OF SEQ ID NOS: 130  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 62  
; LENGTH: 556  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-292-896-62

Query Match 35.9%; Score 1122.5; DB 12; Length 556;  
Best Local Similarity 46.3%; Pred. No. 8e-96;  
Matches 239; Conservative 78; Mismatches 170; Indels 29; Gaps 12;  
Qy 76 GARCEAVRLQCEELRLQESVRLHQLINIVLSDRISLHRLPERWNPCKEKKYDYNL 135  
Db 58 GEMGKAVLIPKDDQ--KMKELFKINQFNLWASDLIALNRSLPDVRLGCKTKYYP-DEL 114  
Qy 136 PRTSVIAFYNEAWSTLLRTVYSVLETSPIILLBEVILVDDYSREHLKERLANELSGLP 195  
Db 115 PNTSVIVFHNEAWSTLLRTVYSVINRSPHYLLSEVILVDDASERDFLKLTVNYKNLE 174  
Qy 196 -KVLIRANKREGLVRALLGASARGVDVLPDCHCECHGWLPLELQRIHEESAVVC 254  
Db 175 VPVKIIRMEERSGLIRARLRGAAASKQVITFLDAHCECTGLWLEPLARIKERKTVC 234  
Qy 255 PVIDVIDMNTFEYLGNSGEPQIGCFDWRVLTWHTVPERERIRMQSPVDV-IRSPTWAGG 313  
Db 235 PVIDVIDMNTFEYLGNSGEPQIGCFDWRVLTWHTVPERERIRMQSPVDV-IRSPTWAGG 293  
Qy 314 LFAVSKKYFEYGLSYDTGMVWGGENLEFSGFRWQCQGVLETHPCSHVGHVFPKQAPYS- 372  
Db 294 LFSIDRNYFEIPTYDAGMDIWMGENLEMSFRIWQCQGSLEIVTCSHVGHVFRKATPTTF 353  
Qy 373 -----RNKALANSVRAAEVWMDPEFKELYHNRNPAARLPFGDVTERRKQLRDLCKQK 425  
Db 354 PGGTGHVINK--NNRRLAEVWMDPEKDFYIISPGVVYKVDYGVSVRKLRENKCKPF 410  
Qy 426 KWFLETYVPELHVPEDEPFGFGLMQLONKGLTDYCFDYNPPDENQIVGHQVILYLCHGMGQ 485  
Db 411 SWYLENIYDPSQIPR-RYVSLGIRNVE--TNQCLDNNMGRKENEKVG----IFNCHGMGN 464



Query Match 35.2%; Score 1100.5; DB 14; Length 559;  
Best Local Similarity 43.0%; Pred.No. 9.2e-94;  
Matches 240; Conservative 86; Mismatches 171; Indels 61; Gaps 17;

18 TACGACAGCAATCCGGTGTTCCTCCTCGTTTAAATGACGCAATGTAATGCCTCATGCC 97

Db	37	KKERGLPAGDV-----LEPYQK-----PHEGPGMGKFPVWPKEQOE---KMKEM 78
Qy	98	VRHQNIYLSDRISLHRRLPERNWPLCKEKKYDYNLPRTSVIIAFNEAWSTLLRTVY 157
Db	79	FKINQFNLMASEMIALNRSUPVRLEGCKTKVP-DNLPTTSVVIVFHNEAWSTLLRTVH 137
Qy	158	SVLETSPODILLEEVLVDDYSREHLKERLANELSGLP-KVRLIRANKREGIVRARLLGA 216
Db	138	SVINRSPRWMLBEIVLVDDASERDFLKRPLESYVKLVKVPVHVIRMEQSRGSLRARLKA 197
Qy	217	SAARGDVLFLDCHCECHGWLBPFLQRIHBEASAVVCPDIVDWNFTFYLGNSGERGEPOI 276
Db	198	AVSGQVITFLDAHCECTVGWLEPFLARIKHDRKTVVCPITDIVSDDTFFYMAGS-DMTY 256
Qy	277	GGFDWRLVFTWHTVPRERIRMQSPVDV-IRSPTMAGGLFAVSKKYFEYGLSGYDTGMVW 335
Db	257	GGFNWKLNFVYVPQREMDRRKGRDTLPVPTPTMAGGLFSIDRDFQEIETYDAGMDIIV 316
Qy	336	GGENLEFSRIWCGGVLETHPCSHVGHVFPFKAQYS-----RNKALANSVRAAEVW 387
Db	317	GGENLEISFRIWCGGTLFIVTCSHVGHVFRKATFYTFPGGTGQIINK--NNRRLAEVW 373
Qy	388	MDEFKELYHNRNPARLEPPGDVTERKQLBDKLOCKDFKMFLETVYVELHVPDRPGPFG 447
Db	374	MDEFKNFFYIISPGVTKVDYGDIISSRLGLRHLKQCRPFSWYLENIYDPSQIPRH---YFS 430
Qy	448	MLQNKGL-TDYCFDYNPPDENQIVGHQVILYLCHGMGNQOFFEYTSOKEIENVNTHQPEGC 506
Db	431	LGBIENVETNQCLJDNWARKENKVG----IFNCHGMGNQVFSYTAKEIKRTD----DLC 482
Qy	507	IAYEAGMDTLIMHLCBETAPENQ-----KFILQEDGSLFHEQSKKCVQQAARKESSDSF 559
Db	483	LDVSKLNGPVTMLKCHH-LKGNQLWEYDPVKLTIQ-----HVNSNQCLDKATDE--DSQ 533
Qy	560	VPLLROCTNSDHQKWFVK 577
Db	534	VPSIRDCSGSRSQOWLLR 551

```

RESULT 15
US-10-205-219-36
; Sequence 36, Application US/10205219
; Publication No. US20030138803A1
; GENERAL INFORMATION:
; APPLICANT: Warner-Lambert Company
; APPLICANT: Lee, Kevin
; APPLICANT: Dixon, Alistair
; APPLICANT: Brooksbank, Robert
; APPLICANT: Pinnock, Robert
; TITLE OF INVENTION: Identification and Use of Molecules Implicated in Pain
; FILE REFERENCE: WL-A-018200
; CURRENT APPLICATION NUMBER: US/10/205,219
; CURRENT FILING DATE: 2002-07-24

```

Job time : 38 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2003, 19:40:38 ; Search time 20 Seconds  
(without alignments)  
2793.700 Million cell updates/sec

Title: US-10-074-527-2

Perfect score: 3124

Sequence: 1 MWGRTARRRCPRRLRGREA.....LLRDCNTSDHQKWFKRMRL 581

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR.76.\*

2: PIR1.\*

3: PIR3.\*

4: PIR4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1194.5	38.2	623	2 T42245	probable polypepti
2	1176	37.6	626	2 T42246	polypeptide N-acet
3	1173	37.5	624	2 T42247	polypeptide N-acet
4	1156.5	37.0	633	2 JC5247	polypeptide N-acet
5	1101.5	35.3	559	2 JC4223	polypeptide N-acet
6	1100.5	35.2	559	2 A45987	polypeptide N-acet
7	1013.5	32.4	571	2 I37405	polypeptide N-acet
8	1002	32.1	563	2 A88515	polypeptide N-acet
9	1002	32.1	612	2 T42243	probable polypepti
10	1001	32.0	589	2 T42244	probable polypepti
11	999	32.0	617	2 T42249	polypeptide N-acet
12	988.5	31.6	579	2 T31549	polypeptide N-acet
13	973.5	31.2	618	2 T42248	polypeptide N-acet
14	949	30.4	562	2 T42250	polypeptide N-acet
15	839.5	26.9	601	2 T42251	polypeptide N-acet
16	808.5	25.9	605	2 T27397	hypothetical prote
17	789.5	25.3	684	2 T26330	hypothetical prote
18	705	22.6	421	2 T42252	polypeptide N-acet
19	299.5	9.6	276	2 T12552	hypothetical prote
20	169.5	5.4	318	2 D87506	glycosyl transfera
21	144.5	4.6	972	2 T09595	glucuronosyltransf
22	142.5	4.6	297	2 D69327	succinoglycan bios
23	141	4.5	313	2 A12404	hypothetical prote
24	141	4.5	749	2 A95953	hypothetical prote
25	138	4.4	312	2 S74669	probable bifunctio
26	135	4.3	309	2 B87550	hypothetical prote
27	132.5	4.2	302	2 A84363	glycosyl transfera
28	128.5	4.1	328	2 B84363	hypothetical prote
29	124	4.0	308	2 H87306	succinoglycan bios

30 122.5 3.9 342 2 G70712  
31 121.5 3.9 755 2 D75598  
32 120 3.8 544 2 B81290  
33 119.5 3.8 334 2 C49348  
34 119.5 3.8 348 2 S39958  
35 119.5 3.8 348 2 D95977  
36 119.5 3.8 348 2 D95977  
37 119.5 3.8 322 2 H69814  
38 118.5 3.8 585 2 B69218  
39 118.5 3.8 294 2 D64175  
40 118 3.8 362 2 G75191  
41 118 3.8 321 2 AG2188  
42 118 3.8 344 2 G70036  
43 118 3.8 358 2 C70171  
44 116 3.7 1044 2 H97186  
45 116 3.7 322 2 AE2160  
46 116 3.7 328 2 F98024

#### ALIGNMENTS

##### RESULT 1

T42245

probable polypeptide N-acetylglactosaminyltransferase (EC 2.4.1.41) - Caenorhabditis el

C;Species: Caenorhabditis elegans

C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jul-2000

C;Accession: T42245

R;Hagen, F.K.; Nehrke, K.

J. Biol. Chem. 273, 8268-8277, 1998

A;Title: cDNA cloning and expression of a family of UDP-N-acetyl-Dgalactosamine:Polypepti

A;Reference number: 222126; MUID:98192620; PMID:9525933

A;Accession: T42245

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-623 <HAG>

A;Cross-references: EMBL:AF031835; NID:G3047190; PIDN:AAC13671.1; PID:G3047191

C;Genetics:

C;Gene: Gly-5

C;Superfamily: polypeptide N-acetylglactosaminyltransferase

C;Keywords: glycosyltransferase; hexosyltransferase

Query Match 38.2%; Score 1194.5; DB 2; Length 623;

Best Local Similarity 45.8%; Pred. No. 5.6e-84;

Matches 257; Conservative 75; Mismatches 178; Indels 51; Gaps 17;

Qy	51	GPRTTRP-----GRREP-----VMPRPV-----PANALGARGAV---RLQ	85
Db	66	GPPIEPVPVNNKVEEPPGCGNLAKPKFWDNDPIYKKGDAQAQAGELGKAVVDKTK	125
Qy	86	LQGEELRLOEESVRLHQINIYLSDRISLHRRLPERNWNPCKEKKYDYDNLPRTSVIATFY	145
Db	126	LSTEEKAKYDKGMNNAFNQYASDMISVHRTLPTNIDAECKTEKYN-ENLPRTSVIICFH	184
Qy	146	NEAMSTLLRTVYSVLTSPDILLEVLVDYSDRHLKRLERLANEUSGL-PKVRLLRANK	204
Db	185	NEAMSVLLRTVHSLVLTSPDILLEVLVDYSDRHLKRLERLANEUSGL-PKVRLLRANK	244
Qy	205	REGVLVARLLGSAARGDVLTFDLCHECEHGLEPLQRIHEESAVCPVIDVDWNT	264
Db	245	REGVLVARLLGSAARGDVLTFDLCHECEHGLEPLQRIHEESAVCPVIDVDWNT	304
Qy	265	FEYLGNSGE-POIIGGFDWRLVFTWHTVPERERIMQSPVDVIRSPMTMAGGLFAVSKKYFE	323
Db	305	FEYHHSKAVFTSVGGFDWGLQFNWHSIPERDRKNRTRPIDVPSPTMAGGLFSIDKEYFE	364
Qy	324	YLGSDYTGMEVWGENLEFSFRIWQCGGVLETHPCSHVGHVFPKQAPY-----SRNKALAN	379
Db	365	KLGTYPGDFDIMGGENLELSFKIWMCGGTLEIVPCSHVGHVFPKRWRTGKRVNLKRN	424
Qy	380	SVPAAEVWDFEKLKYHNSPRARLEPFGDVTERRKOLRDKLCKDFKFWLETYVPELHVP	439
Db	425	SIRLAELWLDYKTYTYIER-INNLQDGFDISRRKKLRDLGCKSKFWYLDNIYPELVFP	483

Qy	440	EDRPGFGLQNGKLTDTYCFDYNPDENQIVGHQVILYLCHGGMGQNOFFYTSQKEIRYN	499
Db	484	GESVAKGEMRNAGGKRNRCIDYKPSG-----GKTVMYQCHNQGNQYMWLSKDGIR--	536
Qy	500	THQPEGCIAVEAGMTLIMHLCETAPENQKFIQED-GLSFHQSKKCVQAAKKESSDS	558
Db	537	--RDESCVDY-AGSDVMYFP-C-HGMKGNQEWYNNHDTGRLQHAVSKCLGTMTKDGAKLE	591
Qy	559	FVPLLRDCTNSD-HQKNFFKE	578
Db	592	MVA-----COYDDPYOHWKFKKE	608

## RESULT 2

T42246  
polypeptide N-acetylglactosaminyltransferase (EC 2.4.1.41) - *Caenorhabditis elegans*  
C;Species: *Caenorhabditis elegans*  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jul-2000  
C;Accession: T42246  
R;Hagen, F.K.; Nehrke, K.  
J. Biol. Chem. 273, 8268-8277, 1998  
A;Title: cDNA cloning and expression of a family of UDP-N-acetyl-Dgalactosamine:Polypeptide  
A;Reference number: Z22126; MUID:98192620; PMID:9525933  
A;Accession: T42246  
A;Status: preliminary; translated from GE/EMBL/DBJ

A:Accession: AF031836; translated from: C9orf252/5550  
A:Molecule type: mRNA  
A:Residues: 1-626 <HAG>  
A:Cross-references: EMBL:AF031836; NID:G3047192; PIDN:AAC13672.1; PID:G3047193  
C:Genetics:  
C:Gene: gly-5  
C:Superfamily: polypeptide N-acetylgalactosaminyltransferase  
C:Keywords: glycosyltransferase; hexosyltransferase

	Query Match	37.6%;	Score 1176;	DB 2;	Length 626;	
	Best Local Similarity	45.6%;	Pred. No. 1.5e-82;			
	Matches 257;	Conservative 80;	Mismatches 173;	Indels 54;	Gaps 20;	
Qy	51 GPPRTPRP-----GRRPE-----VMPRPVP-----PANALGARGEAV---RLQ	85				
	:           :           :					
Dd	66 GPPIEPFVVNNKKVEEEQCGGNLAKPKFWVDNDFIYKGDAAQAGELGKAVVDVK	125				
	:           :           :					
Qy	86 LQGSELALQEBSVRHLQHINIYLSDRISILHRLRPERWNPCLCKEKYDYDNLPRTSVIIAFY	145				
	:           :           :					
Dd	126 LSTSEKAYDKGMINNAFNQASDWISVHRTRLPNIDAECCKEYKN--ENIPRTSVIIICFH	184				
	:           :           :					

128	LSISENARIKIDASMEENAFNQIARSSEISVHKRIUFINIDRECKNEIARIN-ENFPAFVIVICII	D5
146	NEAWSTLLRTVYSVLETSPPDILLBEVILVDDYSOREHLKERLANEISGL-PKVRLIRANK	Qy
185	NEAWSVLLRTVHVSLETPDHLLEBEVLVDDFSMDHTRKPLEEYMSQFGGKVILRMEK	D5
205	REGILVRALLGASAAAGDVLTFIDCHCEHGWLEPLLORIHESASAVCPVIDVIDWNT	Qy
245	REGILRILRGAATGEVLTYLDSHCMEGMWPELLDRIKROPTTTVVCVIDVIDDNT	D5
265	FEYLGNSGE-FQIGGFOWRLVFTWHTTVPERERIRMQSPVDVIRPTWAGGLFAVSKKYPE	Qy
305	FEYHVSNAKYCSVCFENCT-OFNHSINERPEKATERIDVYRLETTWAGGLFSTOWVEF	D5

305	FEYHHSKAYFTSVGGFDWGLQFNHHSIPERUKKNIKRPIDPVRSPTWAGGLFSIDAKAIFE	364
324	YLGSYDTGMEVWGGENLEFFSPRIWQCQGVLETHPCSHGVHVPFPKQAPY	379
365	KLGTYPDGFDIWGGENLELSPKIWWCGGTLEIVPCSHGVHVFKRSPYKWTGYNVLKRN	424
380	SVRAAEVWMDSEFKLYHYHNRPARLEPPGVDVTEKQLRDKLQCKDFKWFLETVTPPELHVP	439
425	SIRLAEVWLDYKTYVYVER-INNLQDGFDISRRKKLREDLGCKSFKWYLDNIYPELVFP	483
440	EDRPGFPGMLNKGILT-DYCFD--YNPPDENQIVGHQVILYLCHGMGONQPFETSQKEI	496

484 GESVA - KGEVRNSAVQPARCLDCMVGRHEKKNRPVG - - - - - TYQCHGGGGGNTWMLSKDGEI 538  
D6  
497 RYNTHQPGCCITAVEAGMDTLIMHICEETAPENQFIQED - GSIFHEOSKKCVQARKES 555  
Qy  
539 R - - - - - RDESCVDY - AGSDVMVFP - C - HGWKMGNOEWRYNHDITGRLOHVASORCKGLMTWKDGA 591  
D6

Qy 556 SDSFVPLLRDCTNSD-HQKWFKE 578  
pB 592 KLEMVA---COYDDPYOHWKFE 611

### RESULT 3

polypeptide N-acetylgalactosaminyltransferase (EC 2.4.1.41) - *Caenorhabditis elegans*  
T42247  
C/Species: *Caenorhabditis elegans*  
C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jul-2000  
C/Accession: T42247  
R/Hagen, F.K.; Nehrl, K.  
J. Biol. Chem. 273, 8268-8277, 1998  
A:Title: cDNA cloning and expression of a family of UDP-N-acetyl-Dgalactosamine:Polypeptide  
A:Reference number: 822126; MUID:98192620; PMID:9525933

A;Accession: T42247  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-624 <HAG>  
A;Cross-references: EMBL:AF031837; NID:G3047194; PIDN:AAC13673.1; PID:G3047195  
C;Genetics:  
A;Gene: gly-5  
C;Superfamily: polypeptide N-acetylgalactosaminyltransferase  
C;Keywords: glycosyltransferase; hexosyltransferase

Query Match 37.5%; Score 1173; DB 2; Length 624;  
Best Local Similarity 45.6%; Pred. No. 2.6e-82;  
Matches 256; Conservative 79; Mismatches 177; Indels 50; Gaps 19;

[illegible]

205	REGIVRARLLGASAA	RGVLTFLDCHCECHG	WLEPLLQRIHHEES	AVCPVIDVIDWNT	264
245	REGILRLRGAA	VATGEVLTYLDSHC	CEGMWLEPLLDRI	KRDPTTVCVIDVIDNT	304
265	FEYLGNSGE	-PQIGGPDWRLVFT	WHITTPERERIMQ	SPDVIRISPTWAGGLFAVSKKVEE	323
305	FEYHH-KAYFTS	VGGGDFDGLQFN	WHSIPERDRKNRT	TRPIDVPSRPTWAGGLFSDKVEE	364
324	YLGSDYTGMEV	GGENLEPFSFR	YWCQCGVLETHPC	SHGVHVPFKQAPY	379
365	YLGSDYTGMEV	GGENLEPFSFR	YWCQCGVLETHPC	SHGVHVPFKQAPY	424

[illegible]

```

Db      538  --RDESCVDY--AGSDVMVFP--C-HGMKGNQEWRYNHDTRGLQHAVSQKCLGWTXDKGAKLE 559
QY      559  FVPLLRLDCTNSD-HQKWFFKE 578
      |         | : | | |
Db      593  MVA-----QYDDPYQHWKFE 609

RESULT 4
JC5247
polypeptide N-acetylgalactosaminyltransferase (BC 2.4.1.41) T3 - mouse

```

## RESULT 4

polypeptide N-acetylglucosaminyltransferase (EC 2.4.1.41) T3 - mouse JC5247



A;Reference number: JC4223; MUID:96115928; PMID:8690719  
A;Accession: JC4223  
A;Molecule type: mRNA  
A;Residues: 1-559 <MEU>  
A;Experimental source: salivary gland  
A;Note: The authors translated the codon AAT for residue 264 as Asp  
R;White, T.; Bennett, E.P.; Paul, E.; Takio, K.; Srensen, T.; Bonding, N.; Clausen, H.  
J. Biol. Chem. 270, 24156-24165, 1995  
A;Title: Purification and cDNA cloning of a human UDP-GalNAc:polypeptide N-Acetylgalactosyltransferase  
A;Reference number: I37404; MUID:96025800; PMID:7592619  
A;Accession: I37404  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-559 <RES>  
A;Cross-references: EMBL:X85018; NID:G971458; PIDN:CAA59380.1; PID:G971459  
C;Comment: This enzyme catalyzes the initial reaction in O-linked (mucin type) oligosaccharide biosynthesis.  
C;Genetics:  
A;Map position: 18  
C;Superfamily: polypeptide N-acetylgalactosaminyltransferase  
C;Keywords: chromosomal protein; glycoprotein; glycosyltransferase; hexosyltransferase; F;9-28/Domain: transmembrane #status predicted <TM>  
F;29-559/Domain: endoplasmic reticulum lumenal #status predicted <LUM>  
F;95,141,552/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F;117,118,288/Binding site: carbohydrate (Thr) (covalent) #status predicted  
F;119/Binding site: carbohydrate (Ser) (covalent) #status predicted

	Query Match	35.3%	Score 1101.5	DB 2	Length 559
	Best Local Similarity	43.0%	Pred. No. 7.2e-77		
	Matches	240	Conservative	86	Mismatches 171; Indels 61; Gaps 17;
QY	38	RAORGAGAGAAEPGPRTPRGREPPVPRPPVPANALGARGEAVRLQGEELRLEES	97		
DB	37	KKERGLPAGDV-----LEPVQK-----PHEGEGMGKPVVIPEDQE--KWKEM	78		
QY	98	VRHQINIVISDRISLHRRILPERWNPLCKEKYVDNLPRTSVIIAFYNEANSTLRTVY	157		
DB	79	FKINQFNLMASEMIALNRSIPDVRLGCKTKYVP--DNLPTTSWIVFHNENASTLLRTVH	137		
QY	158	SVLETSPPDILLEEVILVDDYSDRHLKRLANELSGLP-KVPLIRANKREGVLVARLLGA	216		
DB	138	SVNRSPPHMEIEHVLVDDASERDFLKRPLSEYVVKLVKPVHVIRMEQRSGLIRALKGA	197		
QY	217	SAARGDVLTFDCHCECHGWLPLELQRIHEESAVVCPVIDWNTPEYLGNSGEPQI	276		
DB	198	AVSKGQVITFLDAECCTGVHLEPLLARIKHDRRTIVVCFIIDVISDITFEYWAGS--DMTY	256		
QY	277	GGFDWRLVFTWHTVPERIRIMQSPDV-IRSPTMAGGLFAVSKKYFEYLGSDYTGMEVW	335		
DB	257	GGFNWKLNRWYVPQREMRRKGRDRLPVRTPTMAGGLFSIDRDYFQEGITGYDAGMDIW	316		
QY	336	GGENLESPRIWQCGGVLETHPCSHVGHVFPQAPYS-----RNKALANSVRAAEVW	387		
DB	317	GGENLEISFRWQCGGTLEIVTCSHVGHVFPKATPTTFPGGTGQIINK---NNRRLEAEW	373		
QY	388	MDFFKELYHYRNPRLRPFPGDVTERRKLRDLCKDFKWFLETVYVELHVPEDRPGFFG	447		
DB	374	MDFFKPPFYIISPGVTVDVGDISRRVGLRHLCKPFSWYLENIYDPSQIPRH---YFS	430		
QY	448	MLQNKGL-TPYCFDYNPDENQIVGHQVILYCHGMGNQFFETYSQKEIRYNTHOPEGC	506		
DB	431	LGEBIRNVETNQCLDNMARKENKVG---IFNCHGMGNQVFSYTANKERTD---DLC	482		
QY	507	IAYEAGMDTLIMHLCETAPENQ-----KFLEQDGLSFHEQSKKCVQAARKESDSF	559		
DB	483	LDVSKLNGPVTMLKCHH-LKGNQLWEYDPVKLTQ-----HVNQNQLDKATEE---DSQ	533		
QY	560	VPLLRDCTNSDHOKWFFK	577		
DB	534	VPSIRDNGSRSSQWLLR	551		

RESULT 6



Science 282, 2012-2018, 1998  
A;Title: Genome sequence of the nematode *C. elegans*: a platform for investigating biology  
A;Reference number: A75000; MUID:99069613; PMID:981916  
A;Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C.ele  
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999, and  
A;Accession: A88515  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-563 <STO>  
A;Cross-references: GB:chr\_III; PIDN:AAA28224.1; PID:g289776; GSPDB:GN00021  
C;Genetics:  
C;Map position: 3  
C;Superfamily: polypeptide N-acetyl galactosaminyltransferase

```
Query Match      32.1%; Score 1002; DB 2; Length 563;
Best Local Similarity 41.0%; Pred. No. 3.4e-69;
Matches 227; Conservative 90; Mismatches 178; Indels 58; Gaps 18;

QY 38 RAQRGAGAGAAERGPPTP-----RGRRE-PVMPRPVPPANAL----- 75
DB 4 RARR-----IEPLPAAQHSDPDAPHQPEKQKQVYVDKETANQLAKLMEQAFGP 57
QY 76 GARGE-AVRLOQGEELRLOESVRLHQINILSDRLSHRRRLPERWNLCKEKKYDYD- 133
DB 58 GYHGOGGTGVTPEDKTKIKERFLENQFNVAASEMISVNRTPDYSRDCRTSGNNLKT 117
QY 134 -NLPTSIIIAFNEAWSTLLRTVYSVLETSPIILLEEVILVDDYSDRHLKERLANELS 192
DB 118 AGPKTSIIIVFNEAWTLLRLTSHVNSPRHLEIEIILVDDKSDRYLVKPLDSYIK 177
QY 193 GLP-KVRLIRANKREGVLRLGASARGDVLFTLDCHCEHGWLEPLQLRIHEESA 251
DB 178 MFPIPIHLVHLENRSGILRLTGSMAKGKILLFLDAHVEVTDGWLPLEVSRVAEDRKR 237
QY 252 VVCPVIDVDMNTFEYLGNSGEPQIGGDMRLVFTWHTVPERIRMQSPDVV-IRSPTM 310
DB 238 VVAPIIDVISDDTFEYV-TASETTWGGFNHNLFRWYAVPKRELNRGRSDRSMPIQPTI 296
QY 311 AGGLFAVSKKYFYLGSYDTGMVWGGENLEFSPRIWQCGVLETHPCSHVGHVFPKQAP 370
DB 297 AGGLFAIDKQFFYDYGSDGMQVWGGENLEISFRVWMCGLSLEIHPCSRVGHVFRKQTP 356
QY 371 YSRNKALA-----NSVRAAEVWMDPEFKELYHNRPRARLEPFGDVTBRKOLRKLCKDF 425
DB 357 YTFPGGTAKVIHNAARTAEVWMDYKAFYKMWPAARNVEAGDVSERKKLRETLOCKSF 416
QY 426 KWLETVPYELHVPEDRPGFFGLMQLNGKLTDCYCFDYNPPDENQIVGHQVILYLCHGMGN 485
DB 417 KWLENLYPEALPADFRS-LGAIIVNR-FTEKCVDTNGKKGQAPGIAQ-----CHGAGGN 470
QY 486 QFFEYTSQKEIRYNTHQPECCIA-----VEAGMDTLIMHLCETAPENOKFILOED---GS 538
DB 471 QAMSLTGKGEIR-----SDDLCLSSGHVYQIGSB-LKLERC-SVSKINVHVFFDDOAGT 524
QY 539 LFHEQSKKCVQAA 551
DB 525 LLHKTKGKCVTGA 537
```

## RESULT 9

T42243  
probable polypeptide N-acetyl galactosaminyltransferase (EC 2.4.1.41) - *Caenorhabditis elegans*  
C;Species: *Caenorhabditis elegans*  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Nov-2000  
C;Accession: T42243; S4913  
R;Hagen, F.K.; Nehrke, K.  
J. Biol. Chem. 273, 8268-8277, 1998  
A;Title: cDNA cloning and expression of a family of UDP-N-acetyl-Dgalactosamine:Polypept  
A;Reference number: Z22126; MUID:98192620; PMID:9525933  
A;Accession: T42243  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-612 <HAG>

A;Cross-references: EMBL:AF031833; NID:g3047186; PIDN:AAAC13669.1; PID:g3047187  
R;Wilson, R.  
submitted to the EMBL Data Library, May 1993  
A;Description: Sequence of the *C. elegans* cosmid ZK688.  
A;Reference number: S4913  
A;Accession: S4913  
A;Molecule type: DNA  
A;Residues: 50-612 <WIL>  
A;Cross-references: EMBL:LI16621; NID:g289775; PIDN:AAA28224.1; PID:g289776  
C;Genetics:  
A;Gene: gly-3  
A;Introns: 101/1; 229/3; 520/3; 570/3  
C;Superfamily: polypeptide N-acetyl galactosaminyltransferase  
C;Keywords: glycosyltransferase; hexosyltransferase

```
Query Match      32.1%; Score 1002; DB 2; Length 612;
Best Local Similarity 41.0%; Pred. No. 3.8e-69;
Matches 227; Conservative 90; Mismatches 178; Indels 58; Gaps 18;

QY 38 RAQRGAGAGAAERGPPTP-----RGRRE-PVMPRPVPPANAL----- 75
DB 53 RARR-----IEPLPAAQHSDPDAPHQPEKQKQVYVDKETANQLAKLMEQAFGP 106
QY 76 GARGE-AVRLOQGEELRLOESVRLHQINILSDRLSHRRRLPERWNLCKEKKYDYD- 133
DB 107 GYHGOGGTGVTPEDKTKIKERFLENQFNVAASEMISVNRTPDYSRDCRTSGNNLKT 166
QY 134 -NLPTSIIIAFNEAWSTLLRTVYSVLETSPIILLEEVILVDDYSDRHLKERLANELS 192
DB 167 AGPKTSIIIVFNEAWTLLRLTSHVNSPRHLEIEIILVDDKSDRYLVKPLDSYIK 226
QY 193 GLP-KVRLIRANKREGVLRLGASARGDVLFTLDCHCEHGWLEPLQLRIHEESA 251
DB 227 MFPIPIHLVHLENRSGILRLTGSMAKGKILLFLDAHVEVTDGWLPLEVSRVAEDRKR 286
QY 252 VVCPVIDVDMNTFEYLGNSGEPQIGGDMRLVFTWHTVPERIRMQSPDVV-IRSPTM 310
DB 287 VVAPIIDVISDDTFEYV-TASETTWGGFNHNLFRWYAVPKRELNRGRSDRSMPIQPTI 345
QY 311 AGGLFAVSKKYFYLGSYDTGMVWGGENLEFSPRIWQCGVLETHPCSHVGHVFPKQAP 370
DB 346 AGGLFAIDKQFFYDYGSDGMQVWGGENLEISFRVWMCGLSLEIHPCSRVGHVFRKQTP 405
QY 371 YSRNKALA-----NSVRAAEVWMDPEFKELYHNRPRARLEPFGDVTBRKOLRKLCKDF 425
DB 406 YTFPGGTAKVIHNAARTAEVWMDYKAFYKMWPAARNVEAGDVSERKKLRETLOCKSF 465
QY 426 KWLETVPYELHVPEDRPGFFGLMQLNGKLTDCYCFDYNPPDENQIVGHQVILYLCHGMGN 485
DB 466 KWLENLYPEALPADFRS-LGAIIVNR-FTEKCVDTNGKKGQAPGIAQ-----CHGAGGN 519
QY 486 QFFEYTSQKEIRYNTHQPECCIA-----VEAGMDTLIMHLCETAPENOKFILOED---GS 538
DB 520 QAMSLTGKGEIR-----SDDLCLSSGHVYQIGSB-LKLERC-SVSKINVHVFFDDOAGT 573
QY 539 LFHEQSKKCVQAA 551
DB 574 LLHKTKGKCVTGA 586
```

## RESULT 10

T42244  
probable polypeptide N-acetyl galactosaminyltransferase (EC 2.4.1.41) - *Caenorhabditis elegans*  
C;Species: *Caenorhabditis elegans*  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jul-2000  
C;Accession: T42244  
R;Hagen, F.K.; Nehrke, K.  
J. Biol. Chem. 273, 8268-8277, 1998  
A;Title: cDNA cloning and expression of a family of UDP-N-acetyl-Dgalactosamine:Polypept  
A;Reference number: Z22126; MUID:98192620; PMID:9525933  
A;Accession: T42244  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA

A;Residues: 1-589 <HAG>  
A;Cross-references: EMBL:AF031834; NID:g3047188; PIDN:AAC13670.1; PID:g3047189  
C;Genetics:  
A;Gene: Gly-4  
C;Superfamily: polypeptide N-acetylglalactosaminyltransferase  
C;Keywords: glycosyltransferase; hexosyltransferase

Query Match 32.0%; Score 1001; DB 2; Length 589;  
Best Local Similarity 38.5%; Pred. No. 4.3e-69;  
Matches 232; Conservative 100; Mismatches 215; Indels 56; Gaps 18;

Qy 11 PRELRGREALVLLALLALAGLSVLRQAQAGAGAAEPG--PPRTPR-----PGR 60  
Db 3 PRMLK--MKTGTVLAVLWLEGLAFY--VQSTSSSLRPPGRHPPPLQLDPLIPQNPQ 58  
Qy 61 REPWMP---PPVPA-----NALGARGEA-----VRLQGLBEIRLQESVRLHQI 103  
Db 59 NDEIRPKSAAPPIPTINLAEDTTIHERTEKDVTKTDFVEKFLNKGWQHOGEDKYKANSF 118  
Qy 104 NIYLSDRISLHRLRPERWNPCKEKKYDYNLPTSIVIIAFYNEAWSTLARTVYSVLETS 163  
Db 119 NQASDALNPTRKIPDSREPOCRDVSQVGMQPTTIIITHNEARSLLRTVPSVFNQS 178  
Qy 164 PDILLEEVILVDDYSREHLKERLANSGLPKVRLIRANKREGLVRARLIGASAARGDV 223  
Db 179 PEELLEIVLVDNSQ-----DVEIGKELAQIRITVLNNQREGLIRSRVKGAVARAPV 234  
Qy 224 LTFDCHCEHEGWLEPLQRIHEESAVVCPVIDIDWNTFEYLGNSGPGPIQGGFWRL 283  
Db 235 LTFDLSHIECNQKWLEPLARIAENKAVAPIIDVINDFNVFVGASADLR--GGFDWTL 293  
Qy 284 VFTWHTVPER--ERTMSPVDVIRSPTMAGLFAVSKYFEYLGSDTGMVWGENLEF 342  
Db 294 VFRFEWNEQURKEHRAHPTAPIRSPTMAGLFAISKEWFNELTGYDLDMEVWGENLEM 353  
Qy 343 SFRIWQCGVLETHPCSHVGHVFPKQAPY-----SRNKALANSVRAAEVWMDPEKELYH 397  
Db 354 SFRVWQCGGLEIMPCSRVGHVFRKHPYTPFGSGNVFQKTRAAEVWMDPEKELYH 413  
Qy 398 RNPRLPEPGDVTERRKOLRDLCKQKPFLETVYVPELVDPDRPGFFQMLQNGKLTIDY 457  
Db 414 NVPSARFVNGDITDRLAIRDLQCKSPKYLENNVYPOLEIPRKTG--KSFQMK--IGNL 470  
Qy 458 CFYDNPDPENQIVGHQVLYLCHGNGONQFEYTSQKEIRYNTHQPEGCIAVEAGMD--T 515  
Db 471 CLDSVARKESAPG----LFGCHGTGGNQEWVF--DQLTKTFKNAISQLCLDFSSNTENKT 525  
Qy 516 LIMHLCETAPENQKFIQEDGSLFHEQSKCQVQAARKESSDSFVPLLRDCTNSDHQKWF 575  
Db 526 VTMVKCENLRPDT--MVVEKNGWL--TQGGKLTIVNQGSGDMLIYGACELNNGAQRWI 581  
Qy 576 FKE 578  
Db 582 FEK 584

RESULT 11  
T42249  
polypeptide N-acetylglalactosaminyltransferase (EC 2.4.1.41) 6b - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jul-2000  
C;Accession: T42249  
J. Biol. Chem. 273, 8268-8277, 1998  
A;Title: cDNA cloning and expression of a family of UDP-N-acetyl-Dgalactosamine:Polypept  
A;Reference number: 222126; MUID:98192620; PMID:9525933  
A;Accession: T42249  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-617 <HAG>  
A;Cross-references: EMBL:AF031839; NID:g3047198; PIDN:AAC13675.1; PID:g3047199  
C;Genetics:  
A;Gene: gly-6

C;Superfamily: polypeptide N-acetylglalactosaminyltransferase  
C;Keywords: glycosyltransferase; hexosyltransferase

Query Match 32.0%; Score 999; DB 2; Length 617;  
Best Local Similarity 42.5%; Pred. No. 6.5e-69;  
Matches 224; Conservative 80; Mismatches 149; Indels 74; Gaps 18;

Qy 71 PANALGARGEAVRLQGLBEIRLQESVRLHQINYLSDRISLHRLRPERWNPCKEKKY 130  
Db 94 PHDDWEGGAGVS--HLPQOQKLADSTFAVNQFNILVSDGISVRSLSPEIRKESCRNMTY 152  
Qy 131 DYDNLPTSIVIIAFYNEAWSTLRTVYSVLETSVLETSVLETSVLETSVLETSVLETS 189  
Db 153 P-DNLPTTSVIIYHNEAYSTLRTVWSVIDRSPKELLKEIILVDDFSDFREFLRYPTLDT 211  
Qy 190 ELSGLP--KVLIRANKREGLVRARLIGASAARGDVLTELDCHCEHEGWLEPLLRHES 248  
Db 212 TLKPLPTDIIKIRSKERVLIRAMWGAQAGDVLTELDCHCEHEGWLEPLLRHES 271  
Qy 249 ESAVVCPVIDIDWNTFEYLGNSG--EPQGGFWRLVFTWHTVP--ERERIRMQSPVDVIR 306  
Db 272 RKAVPCVIDIINDNTFOY--QKGIEMFRGGFNWQFRWYGMPTAMAKQHLDDPTGP 329  
Qy 307 SPTMAGLFAVSKYFEYLGSDYDGMVWGENLEFSFRIWQCGVLETHPCSHVGHVFP 366  
Db 330 SPTMAGLFSINRNYFEELGEYDPGMDIWGENLEMSFRIWQCGRVEILPCSHVGHVFR 389  
Qy 367 KQAPY-----SRNKAL--ANSVRAAEVWMDPEKELYHNPRA--RLPEFGDVTERRKOLRDK 419  
Db 390 KSSPHDPFGKSGSKVLNTLLRVAEVMWMDMGHYFYKIAFOAHMRSSIDVSRVEURKK 449  
Qy 420 LQCKDFKWLLETVPY--LHVPEDRPGFFQMLQNGKLTIDYCFDYNP----PDENQIVGHQ 473  
Db 450 LNCCKSFNWLQNVQDFHLEPTPLDR---FORMSN---SNYCTAFPGDTGPKHRLLGSP 503  
Qy 474 VILYCHGNGONQFEYTSQKEIRYNTHQPEGCIAVEAGMDTILMHLCEETAPENQKFI 533  
Db 504 CTW----GFDLWQLWLYTGDRIRTDHLL---CLSVV----- 533  
Qy 534 QEDGSLFHEQSKCQVQAARKESSDSFVPLLRDCTNSDHQKWFKERM 580  
Db 534 ----QLLHTTSDWKIQ-----LKECAGFDTEYWDKPKI 563

RESULT 12

T31549  
polypeptide N-acetylglalactosaminyltransferase (EC 2.4.1.41) 9 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 01-Dec-2000  
C;Accession: T31549; T42253  
R;Matthews, L.  
submitted to the EMBL Data Library, October 1999  
A;Reference number: 221043  
A;Accession: T31549  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-579 <WIL>  
A;Cross-references: EMBL:AL117202; PIDN:CAB57897.1; CESP:Y47D3A.23  
A;Experimental source: Clone Y47D3A  
J. Biol. Chem. 273, 8268-8277, 1998  
R;Hagen, F.K.; Nehrke, K.  
J. Biol. Chem. 273, 8268-8277, 1998  
A;Title: cDNA cloning and expression of a family of UDP-N-acetyl-Dgalactosamine:Polypept  
A;Reference number: 222126; MUID:98192620; PMID:9525933  
A;Accession: T42253  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-202,'D',204-579 <HAG>  
A;Cross-references: EMBL:AF031843; NID:g3047206; PIDN:AAC13679.1; PID:g3047207  
C;Genetics:  
A;Gene: CESP:Y47D3A.23; gly-9  
A;Introns: 45/1; 179/1; 276/1; 343/3; 373/1; 453/3; 491/3; 533/3  
C;Superfamily: polypeptide N-acetylglalactosaminyltransferase  
C;Keywords: glycosyltransferase; hexosyltransferase





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2003, 19:39:32 ; Search time 17 Seconds  
(without alignments)  
1607.207 Million cell updates/sec

Title: US-10-074-527-2

Perfect score: 3124

Sequence: 1 MWGRTARRRCPRELRRGREA.....LLRDCTNSHQKWFKERM 581

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Length	ID	Description
1	1109	35.5	559 1	PAGT_MOUSE
2	1101.5	35.3	559 1	PAGT_HUMAN
3	1100.5	35.2	559 1	PAGT_BOVIN
4	1097	35.1	559 1	PAGT_PIG
5	1097	35.1	559 1	PAGT_RAT
6	1002	32.1	612 1	PAG3_CAEAL
7	121.5	3.9	755 1	BPHY_DEIRA
8	119.5	3.8	348 1	EXOO_RHIME
9	118.5	3.8	294 1	Y596_HAEIN
10	116	3.7	322 1	Y586_ANASP
11	112.5	3.6	1618 1	NEST_HUMAN
12	112.5	3.6	3038 1	TRIO_HUMAN
13	110	3.5	428 1	GGPP_NEUCR
14	110	3.5	477 1	YXNA_STRLI
15	109.5	3.5	1275 1	RFBC_MYXXA
16	108.5	3.5	909 1	Y4GI_RHISN
17	108	3.5	513 1	YTH1_RHOER
18	105.5	3.4	301 1	AMSB_ERWAM
19	103	3.3	446 1	CPWD_HUMAN
20	102.5	3.3	1001 1	PAGT_HUMAN
21	102	3.3	267 1	DPM1_YEAST
22	102	3.3	1527 1	CAIH_MOUSE
23	99.5	3.2	1919 1	HAPP_RAT
24	99	3.2	250 1	Y868_HAEIN
25	99	3.2	589 1	PHBC_ALCEU
26	98.5	3.2	344 1	YIBD_ECOLI
27	98	3.1	431 1	NOCT_HUMAN
28	96	3.1	419 1	HASA_STRPY
29	96	3.1	488 1	ETB1_MOUSE
30	95.5	3.1	900 1	GGAB_BACSU
31	95	3.0	500 1	U713_HUMAN
32	94.5	3.0	1118 1	CARB_YEAST
33	94	3.0	479 1	VGLC_PRVIF

#### RESULT 1

ID	PAGT_MOUSE	STANDARD;	PRT;	559 AA.
AC	O08912;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Polypeptide N-acetylglucosaminyltransferase (EC 2.4.1.41) (Protein-			
DE	UDP acetylglucosaminyltransferase) (UDP-GalNac:polypeptide, N-			
DE	acetylglucosaminyltransferase) (GalNac-T1) (ppGantase-T1).			
GN	GALNT1.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.			
RX	MEDLINE=97298094; PubMed=9153242;			
RA	Hagen F.K., Ten Hagen K.G., Beres T.M., Balys M.M.,			
RA	Vanwuyckhuys B.C., Tabak L.A.;			
RT	"CDNA cloning and expression of a novel UDP-N-acetyl-D-			
RT	galactosamine:polypeptide N-Acetylglucosaminyltransferase.";			
RL	J. Biol. Chem. 272:13843-13848(1997).			
CC	-I- FUNCTION. THIS PROTEIN CATALYZES THE INITIAL REACTION IN O-LINKED			
CC	OLIGOSACCHARIDE BIOSYNTHESIS, THE TRANSFER OF AN N-ACETYL-D-			
CC	GALACTOSAMINE RESIDUE TO A SERINE OR THREONINE RESIDUE ON THE			
CC	PROTEIN RECEPTOR. PAGT AND PAG4 HAVE DIFFERENT SUBSTRATE			
CC	SPECIFICITY.			
CC	-I- CATALYTIC ACTIVITY: UDP-N-acetyl-D-galactosamine + polypeptide =			
CC	UDP + N-acetyl-D-galactosaminyl-polypeptide.			
CC	-I- COFACTOR: MANGANESE AND CALCIUM (BY SIMILARITY).			
CC	-I- PATHWAY: Glycosylation.			
CC	-I- SUBCELLULAR LOCATION: Type II membrane protein. Golgi.			
CC	-I- TISSUE SPECIFICITY: HIGHER EXPRESSION IN KIDNEY, HEART, SMALL			
CC	INTESTINE AND CERVIX AND TO A LESSER EXTENT IN ALL THE OTHER			
CC	TISSUES TESTED.			
CC	-I- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.			
CC	-I- SIMILARITY: Contains 1 ricin B-type lectin domain.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; U73820; AAB58477.1; --			
DR	MGD; MGI:894693; Galnt1.			
DR	GO; GO:0004653; F:polypeptide N-acetylglucosaminyltransferase. . . ; IDA.			
DR	GO; GO:0006493; P:O-linked glycosylation; IDA.			
DR	InterPro; IPR001173; Glyco trans 2.			
DR	InterPro; IPR001772; Ricin_B lectin.			
DR	Pfam; PF00555; Glycos_transf_2; 1.			
DR	Pfam; PF00652; Ricin B lectin; 3.			
DR	SMART; SM00458; RICIN; 1.			





FT CARBOHYD 118 118 O-LINKED (POTENTIAL).  
FT CARBOHYD 119 119 O-LINKED (POTENTIAL).  
FT CARBOHYD 141 141 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 288 288 O-LINKED (POTENTIAL).  
FT CARBOHYD 541 541 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 552 552 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 559 AA; 64219 MW; CD68118CB201EE5B CRC64;  
  
Query Match 35.3%; Score 1101.5; DB 1; Length 559;  
Best Local Similarity 43.0%; Pred. No. 1.9e-77;  
Matches 240; Conservative 86; Mismatches 171; Indels 61; Gaps 17;  
  
Qy 38 RAORGAGAGAAEPGPRTFRGRRPVPVPANALGARGEAVRLQLOGEELRLOEES 97  
Db 37 KKERGLPAGDV-----LEPVQK-----PHEGPGMGKVPVVPKEDQE--KMKEM 78  
  
Qy 98 VRLHQINIVLSDRISLHRLPERWNPCKEKKYDYNLPRTSVIIAFNEAWSTLLRTVY 157  
Db 79 FKINQFNLMASEMIALNRSFDPVLEGCKTKVP-DNLPTTSVIVFNEAWSTLLRTVH 137  
  
Qy 158 SVLETSDDLLEEVILVDDYSRDLKRLERLANELSGLP-KVRLIRANKREGLVARLLGA 216  
Db 138 SVINRSPRHMEETVVLVDASERDFLKRPLESYVKLKVPHVIRMEQORSGLIRALKGA 197  
  
Qy 217 SAARGDVLTFDCHCEHGLELQRIHEESAVVCPVIDVDWNTFEVLGNSGPQI 276  
Db 198 AVSKGQVITFLDAHCECTVGLPELRLARIKHRTVVCPIIDVSDDTFEYMAGS-DMTY 256  
  
Qy 277 GGFOWRLVFTWHTVPERIRIMQSPVDV-IRSPTMAGGLFAVSKKYFEYLGSDYTGMEVW 335  
Db 257 GGFNKLNFWRYPVQREMDRKGRDRLVPTPTMAGGLFSIDRDYQIEITYDAGMDIW 316  
  
Qy 336 GGENLEFSRIWQCGGVLETHPCSHVGHVFPKQAPYS-----RNKALANSVRAAEVW 387  
Db 317 GGENLEFSRIWQCGGVLEIVTCSHVGHVFPKATPTFPGGTGOINK--NNRRLAEVW 373  
  
Qy 388 MDEKELYHHRNPARLEPFGDVTERRKQDLKLOCKFKWPLEVYVELHVPEDRPGFFG 447  
Db 374 MDEPKNFYIISPGTVKVDYGDISRSVRLHKLKQPFSSWYLENIYDPSQIPRH---YFS 430  
  
Qy 448 MLQNKGL-TDYCFDYNPDENOIVGHQVILYLCHGMGNQFPFVTSOKEIRYNTHQEGC 506  
Db 431 LGEIRNVETNQLDNMARKENKVG-----IFNCHGMGNQVFSITANKERTD---DLC 482  
  
Qy 507 IAVEAGMDTLIMHLCEETAPENQ-----KFILQEDGSLFPHQSKVCQVQARKESDSE 559  
Db 483 LDVSKLGPVMTLKCHH-LKGNQLWEYDVPVKLTJQ-----HVNSNQCLDKATEE--DSQ 533  
  
Qy 560 VPLLRLDCTNSDHQKWFPK 577  
Db 534 VPSIRDCNGRSRQOQLLR 551

## RESULT 3

PACT\_BOVIN STANDARD; PRT; 559 AA.  
AC Q07537;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Polypeptide N-acetylgalactosaminyltransferase (EC 2.4.1.41) (Protein-  
DE UDP acetylgalactosaminyltransferase) (UDP-GalNAC:polypeptide, N-  
DE acetylgalactosaminyltransferase) (GalNAC-T1).  
GN GALNT1.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN (1)  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 41-74.  
RC TISSUE=Colostrum, and Intestine;  
RX MEDLINE=93286099; PubMed=7685345;

RA Hona F.L., Hollander T., Lehman D.J., Thomsen D.R., Elhammer A.P.;  
RT "Isolation and expression of a cDNA clone encoding a bovine UDP-  
RL GalNAC:polypeptide N-acetylgalactosaminyltransferase."; J. Biol. Chem. 268:12609-12616(1993).  
RN (2)  
RP SEQUENCE OF 41-559 FROM N.A., AND PARTIAL SEQUENCE.  
RC TISSUE=Colostrum, and Placenta;  
RX MEDLINE=93366815; PubMed=8360184;  
RA Hagen F.K., van Wyckhuysen B., Tabak L.A.;  
RT "Purification, cloning, and expression of a bovine UDP-GalNAC:  
RL polypeptide N-acetyl-galactosaminyltransferase."; J. Biol. Chem. 268:18960-18965(1993).  
CC -!- FUNCTION: THIS PROTEIN CATALYZES THE INITIAL REACTION IN O-LINKED  
CC OLIGOSACCHARIDE BIOSYNTHESIS, THE TRANSFER OF AN N-ACETYL-D-  
CC GALACTOSAMINE RESIDUE TO A SERINE OR THREONINE RESIDUE ON THE  
CC PROTEIN RECEPTOR. THE ENZYME MAY PREFERENTIALLY TRANSFER TO  
CC THREONINE.  
CC -!- CATALYTIC ACTIVITY: UDP-N-acetyl-D-galactosamine + polypeptide =  
CC UDP + N-acetyl-D-galactosaminyl-polypeptide.  
CC -!- COFACTOR: MANGANESE AND CALCIUM.  
CC -!- PATHWAY: Glycosylation.  
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi.  
CC -!- TISSUE SPECIFICITY: COLOSTRUM CONTAINS A SOLUBLE FORM.  
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.  
CC -!- SIMILARITY: Contains 1 ricin B-type lectin domain.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; L07780; AAA30532.1; -;  
DR PIR; A45987; AAA68489.1; -;  
DR InterPro; IPR001173; Glyco\_trans\_2.  
DR InterPro; IPR000772; Ricin B lectin.  
DR Pfam; PF00535; Glycos\_transf\_2; 1.  
DR Pfam; PF00652; Ricin\_B\_lectin; 3.  
DR SMART; SM00458; RICIN; 1.  
DR PROSITE; PS00231; RICIN B LECTIN; 1.  
DR Transferrase; Glycosyltransferase; Transmembrane; Signal-anchor;  
KW Golgi stack; Glycoprotein; Manganese; Calcium; Lectin.  
FT PROPEP 1 40 REMOVED IN SOLUBLE POLYPEPTIDE  
FT CHAIN 41 559 POLYPEPTIDE N-  
FT ACETYL GALACTOSAMINYLTRANSFERASE.  
FT FORM.  
FT DOMAIN 1 8 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 9 28 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
(POTENTIAL).  
FT DOMAIN 29 559 LUMENAL, CATALYTIC (POTENTIAL).  
FT CARBOHYD 439 559 RICIN B-TYPE LECTIN.  
FT CARBOHYD 95 95 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 117 117 O-LINKED (POTENTIAL).  
FT CARBOHYD 118 118 O-LINKED (POTENTIAL).  
FT CARBOHYD 119 119 O-LINKED (POTENTIAL).  
FT CARBOHYD 141 141 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 288 288 O-LINKED (POTENTIAL).  
FT CARBOHYD 552 552 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 559 AA; 64192 MW; E3E538C4DE569B40 CRC64;  
  
Query Match 35.2%; Score 1100.5; DB 1; Length 559;  
Best Local Similarity 43.0%; Pred. No. 2.3e-77;  
Matches 240; Conservative 86; Mismatches 171; Indels 61; Gaps 17;  
  
Qy 38 RAORGAGAGAAEPGPRTFRGRRPVPVPANALGARGEAVRLQLOGEELRLOEES 97  
Db 37 KKERGLPAGDV-----LEPVQK-----PHEGPGMGKVPVVPKEDQE--KMKEM 78  
  
Qy 98 VRLHQINIVLSDRISLHRLPERWNPCKEKKYDYNLPRTSVIIAFNEAWSTLLRTVY 157

```
Db 79 FKIQNLMASEMTALNRLSPDVRLEGCKTKVYP-DNLPTTSVVVFHNEAWSTLLRTVH 137
Qy 158 SVLETSDDLLEEVILVDDYSDRHLKRLANLGLSP-KVRLIRANKREGVLRLARLGA 216
Db 138 SVINRSPRMLLEETVLDVDDASERDLKRLPLESYVKLKVPHVIRMEQSLIRARLGA 197
Qy 217 SAARGDVLTFDCHCEHGWLELQRIHEEBSAVVCPVDVTDVMTFFVLGNSGSPQI 276
Db 198 AVSKGQVITFDLHAECCTVGLWLELARIKHDRKTVVCPIDVLSDDTFEYMAGS-DMTY 256
Qy 277 GGFQRLVFTVHTVPERIRMQSPVDV-IRSPTMAGGLFAVSKKYFEYLGSDYTGMEVW 335
Db 257 GGFNKLNFWRVYVPQREMDRKRKDRTLVPTPTMAGGLFSDIRDRYQIEGTVDAGMDIW 316
Qy 336 GGENLEFSFRWQCGGVLETHPCSHVGHVFPKQAPYS-----RNKALANSVRAAEVW 387
Db 317 GGENLEFSFRWQCGGVLETHPCSHVGHVFPKATPYTPPGTGTQIINK---NNRRLAEVW 373
Qy 388 MDEPKELYHNRPRARLEPFGDVTERKQDRKLOCKDKWPLETVYPHLPVDPDRPGFG 447
Db 374 MDEFKNFYIISPGTVKDYGDISRLRLHKLQCRPFSWLENIYPSQIPRH---YFS 430
Qy 448 MLQNKGL-TDYCFDYNPPDENQIVGHQVILYLCHGMGNQPFYTSOKERYNTHQPEGC 506
Db 431 LGEIRNVTNQCLDNMARKENKVG----IFNCHGMGNQVFSYANKIERTD---DLC 482
Qy 507 IAVBAGMDTLIMHLCETAPENQ-----KFILQEDGSLFHEQSKKCVQAKRESDSF 559
Db 483 LDVSKLNGPVTMLKCHH-LKGNQLWEYDVPVLTQ-----HVNNSQCLDKATDE--DSQ 533
Qy 560 VPLARDCTNSDHQKWF 577
Db 534 VPSIRDCSGRSQQWLLR 551
```

## RESULT 4

```
PAGT_PIG
ID PAGT_PIG STANDARD; PRT; 559 AA.
AC Q29121;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Polypeptide N-acetylglucosaminyltransferase (EC 2.4.1.41) (Protein-
DE UDP acetylglucosaminyltransferase) (UDP-GalNAC:polypeptide, N-
DE acetylglucosaminyltransferase) (GalNAC-TL).
GN GALNT1.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=96318021; PubMed=8748160;
RA Yoshida A., Hara T., Ikenaga H., Takeuchi M.;
RL Glycoconj. J. 12:824-828 (1995).
CC -!- FUNCTION: THIS PROTEIN CATALYZES THE INITIAL REACTION IN O-LINKED
CC OLIGOSACCHARIDE BIOSYNTHESIS, THE TRANSFER OF AN N-ACETYL-D-
CC GALACTOSAMINE RESIDUE TO A SERINE OR THREONINE RESIDUE ON THE
CC PROTEIN RECEPTOR.
CC -!- CATALYTIC ACTIVITY: UDP-N-acetyl-D-galactosamine + polypeptide =
CC UDP + N-acetyl-D-galactosaminyl-polypeptide.
CC -!- COFACTOR: MANGANESE AND CALCIUM (BY SIMILARITY).
CC -!- PATHWAY: Glycosylation.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi.
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC -!- SIMILARITY: Contains 1 ricin B-type lectin domain.
CC -----
```

This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```
CC EMBL; D85389; BAA12800.1; -----
CC DR InterPro; IPR001173; Glyco_trans_2.
CC DR InterPro; IPR000772; Ricin_B_lectin.
CC DR Pfam; PF00535; Glycos_transf_2; 1.
CC DR Pfam; PF00652; Ricin_B_lectin; 3.
CC DR SMART; SM00458; RICIN_1.
CC DR PROSITE; PS0231; RICIN_B_LLECTIN; 1.
CC KW Transferase; Glycosyltransferase; Transmembrane; Signal-anchor;
CC Golgi stack; Glycoprotein; Manganese; Calcium; Lectin.
CC REMOVED IN SOLUBLE POLYPEPTIDE N-
CC ACETYLGLACTOSAMINYLTRANSFERASE (BY
CC SIMILARITY).
CC FT CHAIN 41 559
CC FT SOLUBLE POLYPEPTIDE N-
CC ACETYLGLACTOSAMINYLTRANSFERASE (BY
CC SIMILARITY).
CC FT DOMAIN 1 8
CC FT CYTOPLASMIC (POTENTIAL).
CC FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
CC (POTENTIAL).
CC FT LUMENAL, CATALYTIC (POTENTIAL).
CC FT RICIN B-TYPE LECTIN.
CC FT N-LINKED (GLCNAC...) (POTENTIAL).
CC FT O-LINKED (POTENTIAL).
CC FT CARBOHYD 117 117
CC FT CARBOHYD 118 118
CC FT CARBOHYD 119 119
CC FT CARBOHYD 141 141
CC FT CARBOHYD 288 288
CC FT CARBOHYD 552 552
CC FT N-LINKED (GLCNAC...) (POTENTIAL).
CC FT N-LINKED (GLCNAC...) (POTENTIAL).
CC FT N-LINKED (GLCNAC...) (POTENTIAL).
CC SQ SEQUENCE 559 AA; 64118 MW; 68C2D261A51684C6 CRC64;
```

Query Match 35.1%; Score 1097; DB 1; Length 559;

Best Local Similarity 41.5%; Pred. No. 4.3e-77;

Matches 243; Conservative 84; Mismatches 175; Indels 84; Gaps 17;

```
Qy 9 RCPRLRRGREALVLLALLALAGLSVLRQAGAGAAEPGRPRPRGRPEVMPRP 68
Db 33 KCKEKGRLPA-----GDVLEPVQKPHGPGEMKP-----VWPK- 69
Qy 69 PVPANALGARGAVRLQOGEELRLOEESVRLHQINILYSDRISLHRLRPERWNPLCKE 128
Db 70 -----EDQDKVMEKFKINQFNLMASEMIALNRLSDVRLGCKTK 109
Qy 129 KYDYDMLPRTSVIIAFYNEAWSTLLRTVSVLETSDDLLEEVILVDDYSDRHLKERLA 188
Db 110 VYP-DNLPTTSVVVFHNEAWSTLLRTVSVINRSPRMLLEEIVLVDASERDFLKRPLE 168
Qy 189 NELSGLP-KVRLIRANKREGVLRLARLIGASAAAGDVLTFDCHCEHGWLEPLQRIHE 247
Db 169 SYVKLKVPHVIRMEQSLIRARLKGAAVSKGVITFDLHAECCTVGLWLEPLARIKH 228
Qy 248 ESASAVCPVIDVDMNTFFYLGNSGEPQIGGFDWRLVFTVHTVPERIRMQSPVDV-IR 306
Db 229 DRKTVVCPIDVLSDDTFEYMAGS-DMTYGGFNWKLNFWRVYVPQREMDRKRKDRTLVPR 287
Qy 307 SPTMAGGLFAVSKKYFEYLGSDYTGMEVWGENLEFSFRWQCGGVLETHPCSHVGHVFP 366
Db 288 TPTMAGGLFSDIRDRYQIEGTVDAGMDIWGENLEFSFRWQCGGVLETHPCSHVGHVFR 347
Qy 367 KQAPYS-----RNKALANSVRAAEVWMDPEKELYHNRPRARLEPFGDVTERKQDR 418
Db 348 KATPYTPPGTGTQIINK---NNRRLAEVWMDPEKELYHNRPRARLEPFGDVTERKQDR 404
Qy 419 KLQCKDFKWFLETVPYHLPVDPDRPGFGLMQLNKLTDYCFDYNPPDENQIVGHQVILYL 478
Db 405 KLQCRPFSWLENIYPSQIPRHYS--LGEIRNVE-TNQCLDNMARKENKVG----IFN 458
Qy 479 CHGMGNQVFSYANKIERTD-----DLCVDVSKLNGPVTMLKCHH-LKGNQLWEYDVPV 531
Db 459 CHGMGNQVFSYANKIERTD-----DLCVDVSKLNGPVTMLKCHH-LKGNQLWEYDVPV 513
```

FT	CARBOHYD	141	141	N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD	288	288	O-LINKED (POTENTIAL) .
FT	CARBOHYD	552	552	N-LINKED (GLCNAC. . .) (POTENTIAL) .
SQ	SEQUENCE	559 AA;	64229 MW;	5E36A95D9422C853 CRC64;
Query Match				
Best Local Similarity 41.1%; Score 1097; DB 1; Length 559;				
Matches 241; Conservative 87; Mismatches 173; Indels 86; Gaps 17;				
Qy	9	RCPRELRGRGREALLVLLALLALAGLGSVLRAQRGAGAGAAEPGPRTPRGRRRPPVMPRP	68	
Db	33	KCEKKERGLPA-----GDVLELVQXPHGPGMGXP-----VVIK-	69	
Qy	69	PVPANALGARGEAVRLOQEEELRLOEESVRLHQINITYLSDRISLHRRLLPERWNPCLCKE	128	
Db	70	-----EDQEKMKEMFKINGFNLMASEMIAFNRSLPDVRLEGCKTK	109	
Qy	129	KYDYDNLPRTSVIAAFNEAWSTLLRTVYSVLETSFDILLAEVLLVDVDSYDREHLKERLA	188	
Db	110	VIP-DSLUFTSVVIVFNEAWSTLLRTVHSVKNRSPRHMEIEIVLVDASERDFLKRPLE	168	
Qy	189	NELSGLP-KVRLIRANKREGVLRAALGASAAAGDVLTFDLDCHCEHEGWLPELQLRIHE	247	
Db	169	SYVKLLKVPVHVIRMEQSGRLIRARKGAASVKGQVITFLDAHCECTVGWLEPLLARIKH	228	
Qy	248	EEASVCPVIDVDWNTFEVLGNSGEPIQIGFDWRLVFTWHTVPERIRINQSPDVV-IR	306	
Db	229	DRRTVVCPIIDVISDDTFEYMAGS-DMTYGGFNWKLNFWRYPVQREMDRKRGDRTLPR	287	
Qy	307	SPTMAGGLPAVSKKYFEYLGSDYDTGMVWGGENLEFSFRIWCCGVLETHPCSHGVHVP	366	
Db	288	TPTMAGGLFSIDRDYQGEITYDAGMDIWGENLEISFRIWCCGTLIEITVTCSHGVHVP	347	
Qy	367	KQAPYS-----RNKALANSVRAAEVVMWDFEKELYHYHRNPRARLEPFGDVTERRQLRD	418	
Db	348	KATPYTFPGGTQIINK--NNRRLEAVMMDFEKNFFYIISPGVTKYDYGDISRRVGLRH	404	
Qy	419	KLOCKPKWELEVYVELVHPDORPFGFMLOKNGL-TDYCFDYNPDENQIVGHQVILY	477	
Db	405	KLOCKPFSWYLEINYPDSQIPRH---YFSIGEIRNVETNOCLDMARKENKSVG----	457	
Qy	478	LCHGMGNQPFYETSQKEIRYNTHQPEGCIATVAGMDTLIMHLCETAPENQ-----K	530	
Db	458	NCHGMGNQVSYTANKEIRTD---DLCLDVSKLNGPVTMLKCHH-LKGNQLMEYDPVK	512	
Qy	531	FILOEGDGLSFPHEQSKVCQVAAKESDVSFVPLLRDCTNSDHOQWFFK	577	
Db	513	LTLLQ-----HVNASNQCLDKATSEE--DSQVFSIRDCDTSRGSQOWLLR	551	

```

RESULT 6
PAG3 CAEEL
ID _PAG3 CAEEL STANDARD; PRT; 612 AA.
AC P34678; Q9U003;
AC 01-FEB-1994 (Rel. 28, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Polypeptide N-acetylglactosaminyltransferase 3 (EC 2.4.1.41)
DE (Protein-UDP acetylglactosaminyltransferase) (UDP-GalNAc:polypeptide,
DE N-acetylglactosaminyltransferase) (GalNAc-T1) (ppGANTase 3).
GN Gly-3 OR ZK688-8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=98192620; PubMed=9525933;
RA Hagen F.K., Nehrke K.;
RT "cDNA cloning and expression of a family of UDP-N-acetyl-D-
RT galactosamine:polypeptide N-acetylglactosaminyltransferase sequence
RT homologs from Caenorhabditis elegans."

```

```

RL J. Biol. Chem. 273:8268-8277(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnson L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
RA Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Suleton J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Woldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RL Nature 368:32-38(1994).
RN [3]
RP REVISIONS.
RA Waterston R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THIS PROTEIN CATALYZES THE INITIAL REACTION IN O-LINKED
CC OLIGOSACCHARIDE BIOSYNTHESIS, THE TRANSFER OF AN N-ACETYL-D-
CC GALACTOSAMINE RESIDUE TO A SERINE OR THREONINE RESIDUE ON THE
CC PROTEIN RECEPTOR.
CC -!- CATALYTIC ACTIVITY: UDP-N-acetyl-D-galactosamine + polypeptide =
CC UDP + N-acetyl-D-galactosaminyl-polypeptide.
CC -!- PATHWAY: Glycosylation.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC -!- SIMILARITY: Contains 1 ricin B-type lectin domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF031833; AAC13669.1; -
DR EMBL; L16621; AAR28224.3; -
DR PIR; T42243; T42243.
DR WormPep; ZK688.8; CE29649.
DR InterPro; IPR001173; Glyco_trans_2.
DR InterPro; IPR000772; Ricin_B_lectin.
DR Pfam; PF00535; Glycos_transf_2; 1.
DR Pfam; PF00652; Ricin_B_lectin; 2.
DR SMART; SM00458; RICIN; 1.
DR PROSITE; PS50231; RICIN_B_LECTIN; 1.
DR Transferrase; Glycosyltransferase; Transmembrane; Signal-anchor;
KW Glycoprotein; Lectin.
FT DOMAIN 1 13 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 14 34 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT DOMAIN 35 612 LUMENAL (POTENTIAL).
FT DOMAIN 478 612 RICIN B-TYPE LECTIN.
FT CARBOHYD 146 146 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 612 AA; 68911 MW; 3031C2FE933F9858 CRC64;

Query Match 32.1%; Score 1002; DB 1; Length 612;
Best Local Similarity 41.0%; Pred. No. 1.1e-69;
Matches 227; Conservative 90; Mismatches 178; Indels 58; Gaps 18;

QY 38 RAQCAGAGACAGPGRPTP-----RGRRE-PVWRPPVPANAL-----75
DB 53 RARR-----TEPLPAAQHSDPDPAHPQKQKQVVDVDTANQLRKLMTQARGP 106
QY 76 GARGE-AVRLQGBELRLQESVRLHQINYLSDRISLHRLPWRMPLCKEKKYDYD- 133

```

## RESULT 7

```

BPHY DEIRA
ID_BPHY DEIRA STANDARD; PRT; 755 AA.
AC Q9RZ44;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Bacteriophytochrome (EC 2.7.3.-) (Phytochrome-like protein).
GN BHPH OR DRA0050.
OS Deinococcus radiodurans.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
RX MEDLINE=20036896; PubMed=10567286;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RA "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1."
RL Science 286:1571-1577(1999).
RN [2]
RP CHARACTERIZATION, AND MUTAGENESIS OF MET-259; CYS-289 AND HIS-260.
RX MEDLINE=20085458; PubMed=10617469;
RA Davis S.J., Venter A.V., Vierstra R.D.;
RA "Bacteriophytochromes: phytochrome-like photoreceptors from
RT nonphotosynthetic eubacteria."
RL Science 286:2517-2520(1999).
CC -!- FUNCTION: PHOTORECEPTOR WHICH EXISTS IN TWO FORMS THAT ARE
CC REVERSIBLY INTERCONVERTIBLE BY LIGHT: THE R FORM THAT ABSORBS
CC MAXIMALLY IN THE RED REGION OF THE SPECTRUM AND THE FR FORM THAT
CC ABSORBS MAXIMALLY IN THE FAR-RED REGION. HAS ALSO A SLIGHT BLUE

```

[illegible]



```
Db 184 -EPSFTPKINEPL 196
RESULT 10
YS86 ANASP
ID _YS86 ANASP STANDARD; PRT; 322 AA.
AC P22639;
DT 01-AUG-1991 (Rel. 19, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Putative glycosyl transferase alr2836 (EC 2.4.1.1).
GN ALR2836.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RA "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120."
RT DNA Res. 8:205-213(2001).
RN [2]
RP SEQUENCE OF 1-131 FROM N.A.
RX MEDLINE=90264305; PubMed=2111805;
RA Holland D., Wolk C.P.;
RA "Identification and characterization of hetA, a gene that acts early
RT in the process of morphological differentiation of heterocysts."
RL J. Bacteriol. 172:3131-3137(1990).
CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP003591; BAB74535.1; -
DR EMBL; AF031959; AAC32401.1; ALT INIT.
DR PIR; AE2160; AE2160.
DR InterPro; IPR001173; Glyco trans 2.
DR Pfam; PF00535; Glycos_transf_2; 1.
DR Hypothetical protein; Transferase; Glycosyltransferase;
KW Complete proteome.
SQ SEQUENCE 322 AA; 36388 MW; 753C2FB59327D968 CRC64;
Query Match 3.7%; Score 116; DB 1; Length 322;
Best Local Similarity 29.2%; Pred. No. 0.14;
Matches 68; Conservative 35; Mismatches 82; Indels 48; Gaps 16;
Qy 137 RTSVIIAFYNEAMSTLLRTVYSLV-ETSPDILLBEVLVDYS---DRE---HLKERLAN 189
Db 2 KISVIISNYNYA-RYLSRAINSVLAQTHSDI---EIVVDGSDNSRDVITQLQEQAPD 57
Qy 190 ELSGLPKVRLIRANKREGVLRLLGASAGDVLTFDCHCEHGWLEPFLQRIHE-- 247
Db 58 KIK-----PIFQANQOGQ--GAFNAGFAAATGEVVAFLD---ADDVKPHKLQRIVEVF 106
Qy 248 EESAVCPV--IDVIDN--TFEVLGNSGEPOIG-----GPDWRLVFTHTTVPER 293
Db 107 QTSVDVGMHHLDIIDGNDKTDIDQASTQG-PKLSIEDLASVILQTNACWCFPTSGLAYRR 165
Qy 294 ERIRMQSPVDVIRSPFWAGGLFAVSKKYFEYLGSDYDTGMEVWGGENLEFSPRI 346
Db 166 EVLEKVPFIDPVKRWIRIADGCIIVCT---AFLGKIKTLQ-----ENLAY-YRI 209
```

```
RESULT 11
NEST HUMAN
ID _NEST HUMAN STANDARD; PRT; 1618 AA.
AC P48681; O00552;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Nestin.
GN NES.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93123384; PubMed=1478958;
RA Dahlstrand J., McKay R.D.G., Zimmerman L.B., Lendahl U.;
RA "Characterization of the human nestin gene reveals a close
RT evolutionary relationship to neurofilaments."
RL J. Cell Sci. 103:589-597(1992).
RN [2]
RP SEQUENCE OF 296-309 FROM N.A.
RX MEDLINE=99117355; PubMed=9917366;
RA Yaworsky P.J., Kappen C.;
RA "Heterogeneity of neural progenitor cells revealed by enhancers in
RT the nestin gene."
RL Dev. Biol. 205:309-321(1999).
CC -1- TISSUE SPECIFICITY: CNS STEM CELLS
CC -1- DEVELOPMENTAL STAGE: UPON TERMINAL NEURAL DIFFERENTIATION, NESTIN
CC IS DOWN-REGULATED AND REPLACED BY NEUROFILAMENTS.
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X65964; CAA46780.1; -
DR EMBL; AF004335; AAB64426.1; -
DR PIR; S21424; S21424.
DR Genew; HGNC:7756; NES.
DR MIM; 600915; -
DR GO; GO:0005882; C:intermediate filament; NAS.
DR GO; GO:0007417; P:central nervous system development; NAS.
DR InterPro; IPR001664; IF.
DR Pfam; PF00038; filament; 1.
DR PROSITE; PS00226; IF; 1.
DR Intermediate filament; Coiled coil; Neurone.
KW DOMAIN 1 7 HEAD.
FT DOMAIN 8 312 ROD.
FT DOMAIN 313 1618 TAIL.
FT DOMAIN 8 43 COIL 1A.
FT DOMAIN 44 55 LINKER 1.
FT DOMAIN 56 151 COIL 1B.
FT DOMAIN 152 172 LINKER 12.
FT DOMAIN 173 191 COIL 2A.
FT DOMAIN 192 194 LINKER 2.
FT DOMAIN 195 312 COIL 2B.
SQ SEQUENCE 1618 AA; 176704 MW; C9E9AA48C66534D0 CRC64;
Query Match 3.6%; Score 112.5; DB 1; Length 1618;
Best Local Similarity 19.4%; Pred. No. 2.4;
Matches 122; Conservative 57; Mismatches 207; Indels 243; Gaps 26;
Qy 4 RTARRRCPRELRGRREALLVLLALLALAGLS-----VLRARGAGAGAAEP 50
Db 99 RLARERTTEEVARNRRA--VEAKCARAWLSSQAELEERLEALRVAAHEEVRGLNQA 156
Qy 51 GPPRTPRGRPRPVPVFPANALGARGAVRLQLQGEELRLQEEVRLHQINILYLSDR 110
```









Search completed: December 4, 2003, 19:42:50  
Job time : 19 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: December 4, 2003, 19:40:03 ; Search time 39 Seconds  
(without alignments)  
3844.322 Million cell updates/sec

Title: US-10-074-527-2

Perfect score: 3124

Sequence: 1 MWGRTARRCPRELRRGREA.....LLRDCTNSHQKWFKRM 581

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 23:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rvirus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3124	100.0	581	4 Q8IXK2	Q8ixk2 homo sapien
2	3117	99.8	581	4 Q8NGS4	Q8ngs4 homo sapien
3	2708.5	86.7	576	11 Q8BGT9	Q8bgt9 mus musculus
4	1675	52.8	578	11 Q8N832	Q8n832 mus musculus
5	1651	52.8	578	4 Q8N4A0	Q8n4a0 homo sapien
6	1638	52.4	578	4 Q8N208	Q8n208 homo sapien
7	1562	50.0	284	4 Q8H771	Q8h771 homo sapien
8	1502	48.1	272	4 Q86CT9	Q86ct9 homo sapien
9	1198.5	38.4	623	5 Q95ZJ2	Q95zj2 caenorhabdi
10	1194.5	38.2	623	5 Q61391	Q61391 caenorhabdi
11	1190.5	38.1	622	11 Q8C7U7	Q8c7u7 mus musculus
12	1187.5	38.0	608	4 Q8IYH4	Q8iyh4 homo sapien
13	1185.5	37.9	622	4 Q8NC14	Q8nc14 homo sapien
14	1182.5	37.9	622	4 Q9UIV5	Q9uiv5 homo sapien
15	1180	37.8	626	5 Q9U2J8	Q9u2j8 caenorhabdi
16	1177	37.7	624	5 Q95ZJ1	Q95zj1 caenorhabdi

17	1176.5	37.7	622	11 Q8CED2	Q8ced2 mus musculus
18	1176	37.6	626	5 Q61392	Q61392 caenorhabdi
19	1173	37.5	624	5 Q61393	Q61393 caenorhabdi
20	1156.5	37.0	633	11 P70419	P70419 mus musculus
21	1142.5	36.6	633	4 Q14435	Q14435 homo sapien
22	1126	36.0	650	5 Q8MRC9	Q8mrc9 drosophila
23	1126	36.0	650	5 Q9V7T0	Q9v7t0 drosophila
24	1123.5	36.0	556	11 Q8BLE4	Q8ble4 mus musculus
25	1122.5	35.9	516	4 Q96PX0	Q96px0 homo sapien
26	1122.5	35.9	556	4 Q8IUC8	Q8iuc8 homo sapien
27	1122.5	35.9	556	11 Q8CF93	Q8cf93 mus musculus
28	1102	35.3	623	11 Q9QZ16	Q9qz16 mus musculus
29	1094.5	35.0	582	11 Q8BRT3	Q8brt3 mus musculus
30	1089.5	34.9	536	5 Q95T43	Q95t43 drosophila
31	1089.5	34.9	630	5 Q9VMU3	Q9vmu3 drosophila
32	1023	32.7	639	4 Q96C46	Q96c46 homo sapien
33	1023	32.7	639	4 Q8N3T1	Q8n3t1 homo sapien
34	1023	32.7	639	4 Q96DJ5	Q96dj5 homo sapien
35	1013.5	32.4	571	4 Q10471	Q10471 homo sapien
36	1001	32.0	570	11 Q99ME1	Q99me1 mus musculus
37	1001	32.0	589	5 Q61390	Q61390 caenorhabdi
38	999	32.0	617	5 Q61395	Q61395 caenorhabdi
39	997	31.9	443	5 Q9VQ02	Q9vqg2 drosophila
40	992	31.8	603	11 Q925R7	Q925r7 rattus norv
41	991.5	31.7	603	4 Q8IXJ2	Q8ixj2 homo sapien
42	991	31.7	570	11 Q8BL27	Q8bl27 mus musculus
43	990	31.7	526	11 Q922K5	Q922k5 mus musculus
44	989.5	31.7	579	5 Q61398	Q61398 caenorhabdi
45	988.5	31.6	579	5 Q9U2C4	Q9u2c4 caenorhabdi

## ALIGNMENTS

RESULT 1.

Q8IXK2 ID Q8IXK2 PRELIMINARY; PRT; 581 AA.

AC Q8IXK2; 01-MAR-2003 (TrEMBLrel. 23, Created)

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE UDP-GalNAC-transferase 12.

GN GALNT12.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Bennett E.P.;

RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ132365; CAC80100.2; -

KW Transferase.

SQ SEQUENCE 581 AA; 66938 MW; 8C001D58E103A523 CRC64;

Query Match 100.0%; Score 3124; DB 4; Length 581;  
Best Local Similarity 100.0%; Pred. No. 3.6e-246;  
Matches 581; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MWGRTARRCPRELRRGREALVLLALLAGLGSVLRAGAGAGAAEPGPRTPRGR 60

Db 1 MWGRTARRCPRELRRGREALVLLALLAGLGSVLRAGAGAGAAEPGPRTPRGR 60

Qy 61 REVPMRPPVPNALGARGENVRLQGEELRQEEVRLHQNIVLSDRISLHRLPER 120

Db 61 REVPMRPPVPNALGARGENVRLQGEELRQEEVRLHQNIVLSDRISLHRLPER 120

Qy 121 WNPCLCKEKKYDYNLPRTSVIIAFYNEAWSTLLRTVYVLETSPIIDLEEVILVDDYSR 180

Db 121 WNPCLCKEKKYDYNLPRTSVIIAFYNEAWSTLLRTVYVLETSPIIDLEEVILVDDYSR 180

Qy 181 EHLKRLANELSLGPKVRLFRANKREGLVRALLGSAARGDVLTFDCHCEHGLEP 240

Db 181 EHLKRLANELSLGPKVRLFRANKREGLVRALLGSAARGDVLTFDCHCEHGLEP 240

```

Db 181 EHLKERLANELSLGPKVRLIRANKREGLVRARLLGASAARGDVLTLFDLCHCEHGWLEP 240
Qy 241 LLQRIHEESAVCPVIDVDWNTFEYLGNSGEPQIGGFDWRLVFTWHTVPERERIRMOS 300
Db 241 LLQRIHEESAVCPVIDVDWNTFEYLGNSGEPQIGGFDWRLVFTWHTVPERERIRMOS 300
Qy 301 PVDVIRSPMTAGGLFAVSKYKFEYLGSDYDTGMEVWGGENLEFSFRIWQCGGVLETHPCSH 360
Db 301 PVDVIRSPMTAGGLFAVSKYKFEYLGSDYDTGMEVWGGENLEFSFRIWQCGGVLETHPCSH 360
Qy 361 VGHVFPKQAPYSRNKALANSVRAAEVWMDSEFKELYHNRNPRARLEPGDVTTERKQLRDKL 420
Db 361 VGHVFPKQAPYSRNKALANSVRAAEVWMDSEFKELYHNRNPRARLEPGDVTTERKQLRDKL 420
Qy 421 QCKDFKFWLETVPYVPELHVPEDRPGFQGLNQKGLTDYCFDYNPPDENQIVGHQVILYLCH 480
Db 421 QCKDFKFWLETVPYVPELHVPEDRPGFQGLNQKGLTDYCFDYNPPDENQIVGHQVILYLCH 480
Qy 481 GMGNQFFEYTSQKEIRYNTHQPEGCIATVEAGMDTLIMHLCETAPENQKFIQEDGSLF 540
Db 481 GMGNQFFEYTSQKEIRYNTHQPEGCIATVEAGMDTLIMHLCETAPENQKFIQEDGSLF 540
Qy 541 HEQSKKCQVQAARKESSDSFVPLLDRDCTNSDHQKWFKKERML 581
Db 541 HEQSKKCQVQAARKESSDSFVPLLDRDCTNSDHQKWFKKERML 581

```

## RESULT 2

```

Q8NG54 ID Q8NG54 PRELIMINARY; PRT; 581 AA.
AC Q8NG54;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE UDP-N-acetyl-alpha-D-galactosamine:polypeptide
DE N-acetylgalactosaminyltransferase 12.
GN GALNT12.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=22131447; PubMed=12135769;
RA Guo J.M., Zhang Y., Cheng L., Iwasaki H., Wang H., Kubota T.,
RA Tachibana K., Narimatsu H.;
RT "Molecular cloning and characterization of a novel member of the UDP-
RT Gainac:polypeptide N-acetylgalactosaminyltransferase family, pp-
RT Gainac-T12(1).";
RL FEBS Lett. 524:211-218(2002).
DR EMBL; AB078146; BAC07181.1; -.
DR Genbank; HGNC:19877; GALNT12.
DR InterPro; IPR001173; Glyco trans. 2.
DR InterPro; IPR000772; Ricin B lectin.
DR Pfam; PF00535; Glycos transf_2; 1.
DR SMART; SM00458; RICIN; 1.
DR PROSITE; PS50231; RICIN_B_LECTIN; 1.
SQ SEQUENCE 581 AA; 66908 MW; D02D1B2199508FE6 CRC64;

```

```

Query Match 99.8%; Score 3117; DB 4; Length 581;
Best Local Similarity 99.8%; Pred. No. 1.4e-245;
Matches 580; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MWGTARRRCPRELRGREALLVLLALLAGLGSVLRAQAGAGAAEPGPPTRPGR 60
Db 1 MWGTARRRCPRELRGREALLVLLALLAGLGSVLRAQAGAGAAEPGPPTRPGR 60
Qy 61 REPWMPRPVPPANALGARGEAVALRQLOGEELRQESVRLHQINILYSDRISLHRLPER 120
Db 61 REPWMPRPVPPANALGARGEAVALRQLOGEELRQESVRLHQINILYSDRISLHRLPER 120
Qy 121 WNPICKCKKKYDYNLPRTSVIIAFYNEAWSTLLRTVYVLETSPIILLVEVILVDDYSDR 180

```

```

Db 121 WNPICKCKKKYDYNLPRTSVIIAFYNEAWSTLLRTVYVLETSPIILLVEVILVDDYSDR 180
Qy 181 EHLKERLANELSLGPKVRLIRANKREGLVRARLLGASAARGDVLTLFDLCHCEHGWLEP 240
Db 181 EHLKERLANELSLGPKVRLIRANKREGLVRARLLGASAARGDVLTLFDLCHCEHGWLEP 240
Qy 241 LLQRIHEESAVCPVIDVDWNTFEYLGNSGEPQIGGFDWRLVFTWHTVPERERIRMOS 300
Db 241 LLQRIHEESAVCPVIDVDWNTFEYLGNSGEPQIGGFDWRLVFTWHTVPERERIRMOS 300
Qy 301 PVDVIRSPMTAGGLFAVSKYKFEYLGSDYDTGMEVWGGENLEFSFRIWQCGGVLETHPCSH 360
Db 301 PVDVIRSPMTAGGLFAVSKYKFEYLGSDYDTGMEVWGGENLEFSFRIWQCGGVLETHPCSH 360
Qy 361 VGHVFPKQAPYSRNKALANSVRAAEVWMDSEFKELYHNRNPRARLEPGDVTTERKQLRDKL 420
Db 361 VGHVFPKQAPYSRNKALANSVRAAEVWMDSEFKELYHNRNPRARLEPGDVTTERKQLRDKL 420
Qy 421 QCKDFKFWLETVPYVPELHVPEDRPGFQGLNQKGLTDYCFDYNPPDENQIVGHQVILYLCH 480
Db 421 QCKDFKFWLETVPYVPELHVPEDRPGFQGLNQKGLTDYCFDYNPPDENQIVGHQVILYLCH 480
Qy 481 GMGNQFFEYTSQKEIRYNTHQPEGCIATVEAGMDTLIMHLCETAPENQKFIQEDGSLF 540
Db 481 GMGNQFFEYTSQKEIRYNTHQPEGCIATVEAGMDTLIMHLCETAPENQKFIQEDGSLF 540
Qy 541 HEQSKKCQVQAARKESSDSFVPLLDRDCTNSDHQKWFKKERML 581
Db 541 HEQSKKCQVQAARKESSDSFVPLLDRDCTNSDHQKWFKKERML 581

```

## RESULT 3

```

Q8BGT9 ID Q8BGT9 PRELIMINARY; PRT; 576 AA.
AC Q8BGT9;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical glycosyl transferase.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cecum, and Thymus;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK03638; BAC28401.1; -.
DR EMBL; AK042133; BAC31179.1; -.
KW Hypothetical protein.
SQ SEQUENCE 576 AA; 66540 MW; 706310A99FCDACD5 CRC64;

```

```

Query Match 86.7%; Score 2708.5; DB 11; Length 576;
Best Local Similarity 85.2%; Pred. No. 2.8e-212;
Matches 494; Conservative 41; Mismatches 40; Indels 5; Gaps 2;

```

```

Qy 1 MWGTARRRCPRELRGREALLVLLALLAGLGSVLRAQAGAGAAEPGPPTRPGR 60
Db 1 MWGTARRRCPRELRGREALLVLLALLAGLGSVLRAQAGAGAAEPGPPTRPGR 55
Qy 61 REPWMPRPVPPANALGARGEAVALRQLOGEELRQESVRLHQINILYSDRISLHRLPER 120
Db 56 HEPVLPVPPADALGAHGAVALRQLOGEELRQESVRLHQINILYSDRISLHRLPER 115
Qy 121 WNPICKCKKKYDYNLPRTSVIIAFYNEAWSTLLRTVYVLETSPIILLVEVILVDDYSDR 180
Db 116 WNPICKCKKKYDYNLPRTSVIIAFYNEAWSTLLRTVYVLETSPIILLVEVILVDDYSDR 175

```

QY 181 EHLKERLANELSGLPKVLRLTRANKREGVRLARLLGASAGRDVLTFLDCHCEHGWLEP 240  
DB 176 EHLKERLANELSGLPKVLRLTRANKREGVRLARLLGASAGRDVLTFLDCHCEHGWLEP 235  
QY 241 LLQRIHEEESAVCPVLDVTDWNTFEYLSNGSGEPQIGGFDWRLVFTWHTVPERIRMQS 300  
DB 236 LLQRIHEEESAVCPVLDVTDWNTFEYLSNGSGEPQIGGFDWRLVFTWHTVPERIRMQS 295  
QY 301 PVDVIRPTWAGGLFAVSKYKFEYLSYDTCMEYMGGENLEFSPRIWQCGVLETHPCSH 360  
DB 296 PIDVIRPTWAGGLFAVSKYKFEYLSYDTCMEYMGGENLEFSPRIWQCGVLETHPCSH 355  
QY 361 VGHVFPKQAPYSRNKALANSVRAAEVWMDPEKELYIYRNTPARLEPFGDVTERKQLRDLK 420  
DB 356 VGHVFPKQAPYSRNKALANSVRAAEVWMDPEKELYIYRNTPARLEPFGDVTERKQLRDLK 415  
QY 421 QCKDFKWFLETVPELVHVPEDRPGFQGLMQLNGLTDCYFDPDENQIVGHQVILYLCH 480  
DB 416 QCKDFKWFLETVPELVHVPEDRPGFQGLMQLNGLTDCYFDPDENQIVGHQVILYLCH 475  
QY 481 GMGQNFPEYTSQKEIRYNTHQPEGCIAGEAGMDTLIMHLCEETAPENQKFIQEDGSLF 540  
DB 476 GMGQNFPEYTSQKEIRYNTHQPEGCIAGEAGMDTLIMHLCEETAPENQKFIQEDGSLF 535  
QY 541 HEGSKVCQVQARKESDSEFVPLLRDCTNSDHQKWFPERM 580  
DB 536 HEGSKVCQVQARKESDSEFVPLLRDCTNSDHQKWFPERM 575

## RESULT 4

O08832 ID O08832 PRELIMINARY; PRT; 578 AA.  
AC O08832;  
DT 01-JUL-1997 (TrEMBLrel. 04, Created)  
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Polypeptide N-acetylglucosaminyltransferase (EC 2.4.1.41) (Protein-  
DE UDP acetylglucosaminyltransferase) (UDP-GalNac:polypeptide, N-  
DE acetylglucosaminyltransferase) (GalNac-T4).  
GN GALNT4 OR PGCANASB-T4.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.  
RX MEDLINE=97298094; PubMed=9153242;  
RA Hagen F.K., Ten Hagen K.G., Beres T.M., Balys M.M.,  
RA VanWuyckhuysen B.C., Tabak L.A.;  
RT "CDNA cloning and expression of a novel UDP-N-acetyl-D-  
RT galactosamine:polypeptide N-Acetylglucosaminyltransferase.";  
RL J. Biol. Chem. 272:13843-13848(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Colon;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium,  
RA "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs.";  
RL Nature 420:563-573(2002).  
CC -1- FUNCTION: THIS PROTEIN CATALYZES THE INITIAL REACTION IN O-LINKED  
CC OLIGOSACCHARIDE BIOSYNTHESIS. THE TRANSFER OF AN N-ACETYL-D-  
CC GALACTOSAMINE RESIDUE TO A SERINE OR THREONINE RESIDUE ON THE  
CC PROTEIN RECEPTOR. AGT AND PAG4 HAVE DIFFERENT SUBSTRATES  
CC SPECIFICITY.  
CC -1- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GALACTOSAMINE + POLYPEPTIDE =  
CC UDP + N-ACETYL-D-GALACTOSAMINYL-POLYPEPTIDE.  
CC -1- PATHWAY: GLYCOSYLATION.  
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI.  
CC -1- TISSUE SPECIFICITY: HIGHER EXPRESSION IN SUBLINGUAL GLAND,  
CC STOMACH, SMALL INTESTINE, CERVIX AND COLON AND TO A LESSER EXTENT

CC IN KIDNEY, OVARY, LUNG AND UTERUS. LOW LEVELS IN SPLEEN AND TRACE  
CC LEVELS IN LIVER, HEART AND BRAIN. NO EXPRESSION IN SUBMANDIBULAR  
CC AND PAROTID GLANDS, SKELETAL MUSCLE AND TESTIS.  
CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.  
DR EMBL: U73819; AAB58301.1; -;  
DR EMBL: AK033494; BAC28317.1; -;  
DR MGD; MGI:894692; Galnt4.  
DR InterPro; IPR001173; Glyco\_trans\_2.  
DR InterPro; IPR000772; Ricin\_B\_lectin.  
DR Pfam; PF00535; Glycos\_transf\_2; 1.  
DR Pfam; PF00652; Ricin\_B\_lectin; 2.  
DR SMART; SM00458; RICIN\_1.  
DR PROSITE; PS00231; RICIN\_B\_LECTIN; 1.  
KW Transferase; Glycosyltransferase; Transmembrane; Signal-anchor;  
KW Golgi stack; Glycoprotein.  
FT DOMAIN 1 12 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 13 31 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).  
FT DOMAIN 32 578 LUMENAL, CATALYTIC (POTENTIAL).  
FT CARBOHYD 471 471 N-LINKED (GLCNAC...) (POTENTIAL).  
SQ SEQUENCE 578 AA; 66554 MW; 10ADC0DB8B30835 CRC64;  
Query Match 53.6%; Score 1675; DB 11; Length 578;  
Best Local Similarity 57.9%; Pred. No. 5.7e-128;  
Matches 330; Conservative 78; Mismatches 144; Indels 18; Gaps 8;  
QY 23 VLIALALAGL-----GSVLRAQAGAGAGAAEPGPRTPRPRGRR-----PVNPRPPVPA 72  
DB 13 LLLALLTLAYILVEFVSSTLYASPGAG-GARELGPRRLPDLDTREEDLSQPLYIKPADS 71  
QY 73 NALGARGEAVRLQOGEELRLOEESVRLHQINILYLSDRISLHRLPERWNPCLKEKYDY 132  
DB 72 HALGEWGRASKLQNEGELKQKEELIERYAINILYLSDRISLHRIEDKRMVECKAKFY 131  
QY 133 DNLPTSTVIAFYNEAWSTLLRTVYSVLETSPPILLEEVILVDVDSYDREHLKERLANELS 192  
DB 132 RSLPTSTVIAFYNEAWSTLLRTVHSVLETSPPAVLLKEILLVDVDSYDRIYLLKQLEYIS 191  
QY 193 GLPKVRLIRANKREGVRLARLLGASAGRDVLTFLDCHCEHGWLEPLQRIHEESAV 252  
DB 192 NLBRVRLIRANKREGVRLARLLGATFATGDLVLTFLDCHCECNCTGWLPLELLERISRDSTA 251  
QY 253 VCPVIDVIDWNTFEYLSNGSGEPQIGGFDWRLVFTWHTVPERIRMQSPVDVIRPTMAG 312  
DB 252 VCPVIDVIDWNTFEYLSNGSGEPQIGGFDWRLVFTWHTVPERIRMQSPVDVIRPTMAG 311  
QY 313 GLFAVSKYFYLGSYDTCMEYMGGENLEFSPRIWQCGVLETHPCSHVGHVFPKQAPYS 372  
DB 312 GLFAVSKYFYLGSYDTCMEYMGGENLEFSPRIWQCGVLETHPCSHVGHVFPKQAPYS 371  
QY 373 RNKALANSVRAAEVWMDPEKELYIYRNTPARLEPFGDVTERKQLRDLKCKFWFLETY 432  
DB 372 RPNFLQNTARAAEVWMDPEKELYIYRNTPARLEPFGDVTERKQLRDLKCKFWFLETY 431  
QY 433 YPELVHVPEDRPGFQGLMQLNGLTDCYFDPDENQIVGHQVILYLCHGQVQOFFEYTS 492  
DB 432 FSNLHVPEDRPGVHGAIRSMGSISSCLDYNADPNNP-TGANLSLFGCHGQVQOFFEYTS 490  
QY 493 QKEIRYNTHQPEGCIAGEAGMDTLIMHLCEE---TAPENQKFIQEDGSLFHEQSCKVCQ 549  
DB 491 NKEIRFNS-VTELCAEVPOOKDVGQNCNCPQGLPVPVNIHFHFKEDGTIFHPTHRLCLS 549  
QY 550 AARKESDSEFVPLLRDCTNSD-HQKWFPERM 578  
DB 550 AYTAGRSPSVH-MKTCDDALDKQLARFEK 578  
RESULT 5  
Q8N4A0 ID Q8N4A0 PRELIMINARY; PRT; 578 AA.  
AC Q8N4A0;  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)





```
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ21212.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon;
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Ohayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isegai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK024865; BAB15027.1; -.
DR InterPro; IPR000772; Ricin B lectin.
DR Pfam; PF00652; Ricin B lectin; 2.
DR SMART; SM00458; RICIN; 1.
DR PROSITE; PS50231; RICIN_B_LECTIN; 1.
KW Hypothetical protein.
SQ SEQUENCE 284 AA; 33037 MW; 3DD5FD6F1154AB71 CRC64;

Query Match 50.8%; Score 1562; DB 4; Length 284;
Best Local Similarity 100.0%; Pred. No. 3.5e-119;
Matches 284; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 298 MQSPVDVIRPTMAGGLPAVSKKYFEYLGSDYTGMEVWGGENLEFSFRWQCGGVLETHP 357
DB 1 MQSPVDVIRPTMAGGLPAVSKKYFEYLGSDYTGMEVWGGENLEFSFRWQCGGVLETHP 60

QY 358 CSHVGHVFPKQAPYSRNKALANSVRAAEVWMDPEFKELYHNRNPRARLEPGDVTFRKQLR 417
DB 61 CSHVGHVFPKQAPYSRNKALANSVRAAEVWMDPEFKELYHNRNPRARLEPGDVTFRKQLR 120

QY 418 DKLOCKDFKFLVTVYPPELVHPEDRPGFFGMQLNKGTLTDYCFDYNPPDENQIVGHQVILY 477
DB 121 DKLOCKDFKFLVTVYPPELVHPEDRPGFFGMQLNKGTLTDYCFDYNPPDENQIVGHQVILY 180

QY 478 LCHGMGQNFPEYTSQKEIRYNTHQPEGCTAVERAGMDTLIMHLCETAPENQKFILOEDG 537
DB 181 LCHGMGQNFPEYTSQKEIRYNTHQPEGCTAVERAGMDTLIMHLCETAPENQKFILOEDG 240

QY 538 SLFHEQSKKCVQAAKESDSFVPLLRDCTNSDHQKWFKERM 581
DB 241 SLFHEQSKKCVQAAKESDSFVPLLRDCTNSDHQKWFKERM 284

RESULT 8
Q96CT9 ID Q96CT9 PRELIMINARY; PRT; 272 AA.
AC Q96CT9
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to hypothetical protein FLJ21212.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RA Straubeberg B.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC013945; AAH13945.1; -.
DR InterPro; IPR000772; Ricin B lectin.
DR Pfam; PF00652; Ricin B lectin; 2.
DR SMART; SM00458; RICIN; 1.
DR PROSITE; PS50231; RICIN_B_LECTIN; 1.
KW Hypothetical protein.
SQ SEQUENCE 272 AA; 31726 MW; 6E78ADD3FC84390F CRC64;
```

```
Query Match 48.1%; Score 1502; DB 4; Length 272;
Best Local Similarity 100.0%; Pred. No. 2.6e-114;
Matches 272; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 310 MAGGLFAVSKKYFEYLGSDYTGMEVWGGENLEFSFRWQCGGVLETHPCHSHVGHVFPKQA 369
DB 1 MAGGLFAVSKKYFEYLGSDYTGMEVWGGENLEFSFRWQCGGVLETHPCHSHVGHVFPKQA 60

QY 370 PYSRNKALANSVRAAEVWMDPEFKELYHNRNPRARLEPGDVTFRKQLRDKLOCKDFKFL 429
DB 61 PYSRNKALANSVRAAEVWMDPEFKELYHNRNPRARLEPGDVTFRKQLRDKLOCKDFKFL 120

QY 430 ETVPPELVHPEDRPGFFGMQLNKGTLTDYCFDYNPPDENQIVGHQVILYLCHEGMDGQNF 489
DB 121 ETVPPELVHPEDRPGFFGMQLNKGTLTDYCFDYNPPDENQIVGHQVILYLCHEGMDGQNF 180

QY 490 YTSQKEIRYNTHQPEGCTAVERAGMDTLIMHLCETAPENQKFILOEDGSLFHEQSKKCVQ 549
DB 181 YTSQKEIRYNTHQPEGCTAVERAGMDTLIMHLCETAPENQKFILOEDGSLFHEQSKKCVQ 240

QY 550 AAKESDSFVPLLRDCTNSDHQKWFKERM 581
DB 241 AAKESDSFVPLLRDCTNSDHQKWFKERM 272

RESULT 9
Q95ZJ2 ID Q95ZJ2 PRELIMINARY; PRT; 623 AA.
AC Q95ZJ2
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Y39E4B.12b protein.
GN Y39E4B.12 OR Y39E4B.12b.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Barlow K.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; AL110487; CAC42368.1; -.
DR WormPep; Y39E4B.12b; CE28119.
DR InterPro; IPR001173; Glyco_trans_2.
DR InterPro; IPR000760; Inositol_P.
DR Pfam; PF00535; Glycos transf_2; 1.
DR Pfam; PF00652; Ricin B lectin; 3.
DR SMART; SM00458; RICIN; 1.
DR PROSITE; PS00630; IMP_2; 1.
DR PROSITE; PS50231; RICIN_B_LECTIN; 1.
SQ SEQUENCE 623 AA; 71013 MW; 722AC7E93EF5FB4D CRC64;

Query Match 38.4%; Score 1198.5; DB 5; Length 623;
Best Local Similarity 46.0%; Pred. No. 4.6e-89;
Matches 258; Conservative 74; Mismatches 178; Indels 51; Gaps 17;

QY 51 GPPRTPRP-----GRREP-----VMPRPV-----PANALGARGEAV---RLQ 85
DB 66 GPPIEPEPVVNNKVVEEQPGGNLAKPKFVDPNDPIYKKGDAQAQAGELGKAVVVDKTK 125

QY 86 LQGEELRQESVRLHQINILYSDRISLHRLRPERWNPLCKEKKYDYNLPRTSVIAFY 145
DB 126 LSTEERAKYDKGMNNAFNQASDMISVHRTLPTNIDAECKTEKYN-ENLPRTSVIICFH 184
```



```
QY 97 SVRLHQINILSDRISLHRL-PERWNPLCCKEKKY-DYDNLPRTSVLIIFNEAWSTLLR 154
Db 136 GYKXKHCNAPASDRISQSLGSDTRPECPPLTTSVLIIFNEAWSTLLR 195
QY 155 TVSVLETSPIILLEEVILVDDSDRHLKERLANELSGPKVRLIRANKREGILVARLL 214
Db 196 TVSVLHTSPAILLKEILLVDDASTDEHLKERLEQYVQQLQIVRVQRERKGLITARLL 255
QY 215 GASARGDVLTFDCHCEHEGWLPLELLQRIHEESAVVCPVIDVDMNTFEY---LGNS 271
Db 256 GASVAQAEVLTFDACHCEPHGWLPLELLARIAEDKTAVVSPDIVTIDLNTFQSRPVRG 315
QY 272 GEQICGFQDLVTFWHTVPERIRMQSPVDVIRSPMTAGGLFAVSKKYFEYLGSYDTG 331
Db 316 KAHSRGNFWSLTFGWEMLPEHEKQRRKDETYPIKSPTFAGGLFSISKAYFEHIGTYDNQ 375
QY 332 MEVWGENLEFSFRIMOCGVLTHPCSHGVHPFKOAPYSRNKALA---NSVRAAEVW 387
Db 376 MEIWGENVMSFRVWQCGQLIIPCSVVGHVFTKSPHTFPKGTSVIARNQVRLAEVW 435
QY 388 MDEFKELYHRNPRA---RLEPFGDVTERKQLDKLQCKDFKWFLETVYPPELHVPEDRP 443
Db 436 MDDYKKIFYRNLOQAAMWQNNFGDISERLRLREQLRCHNFSWYLNHNVPFVPELNP 495
QY 444 GFTGMLQNKGLTDYCDYNPPDENQIVGHQVILYLCHGMQONOFFEYTSOKEIRYNTHQP 503
Db 496 TIFYGAIKNLG-TNQCILDVG---ENNRGGKPLIMVYCHNLGNQYFEYTSQDLRHNIGK 551
QY 504 EGCIAVEAGMDTILMHLCE---ETAPENOKFILOEDGSLFHEQSKKVOAARKSSDS 558
Db 552 ---LCHASGSLGLRSCQFVGKNSRVPKDEEWELTQDLIRNSGGTCLTSQDKK--- 604
QY 559 FVPLLRDCTNSD-HQKWF 576
Db 605 --PAMAPCNPRDPQLWLF 621
```

## RESULT 12

```
Q81YH4
ID Q81YH4 PRELIMINARY; PRT; 608 AA.
AC Q81YH4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE UDP-N-acetyl-alpha-D-galactosamine:polyepitide
DE N-acetyl-galactosaminyltransferase 6 (GalNac-T6).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymph;
RA Strausberg R.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC035822; AAH35822.1; -.
SQ SEQUENCE 608 AA; 69357 MW; A922107DA2E2443F CRC64;
```

Query Match 38.0%; Score 1187.5; DB 4; Length 608;  
Best Local Similarity 45.9%; Pred. No. 3.5e-88;  
Matches 249; Conservative 91; Mismatches 163; Indels 39; Gaps 14;

```
QY 58 PGRPEVMPRPVPPANALGARGEAVR---LQIQGEELRLQESVRLHQINILSDRISL 113
Db 82 PAELKPFWERPPQDPNAPGADGKAFQSKWTPLETOE---KEGYKXHCNAPASDRISL 138
QY 114 HRL-PERWNPLCCKEKKY-DYDNLPRTSVLIIFNEAWSTLLRTVSVLETSPIILLEEV 171
Db 139 QRSGLGDTTPPECVDQKFRCPPLATTSVLIIFNEAWSTLLRTVSVLHTTFAILLKEI 198
QY 172 ILVDDYSDRHLKERLANELSGPKVRLIRANKREGILVARLLGASARGDVLTFDCHC 231
Db 199 ILVDDASTEEHLKERLEQYVQQLQIVRVVRQERKERKGLITARLLGASVAQAEVLTFDACH 258
```

```
QY 232 ECEHGLEPILLQRIHEESAVVCPVIDVDMNTFEYLGNSEGPQI---GGFDWRLVFTWH 288
Db 259 ECFHGLEPILLARIAEDKTAVVSPDIVTIDLNTFEFAKPVQGRVHSRGNFDSLTFGWE 318
QY 289 TVPERIRMQSPVDVIRSPMTAGGLFAVSKKYFEYLGSYDTGMEVWGGNLEFSFRIMQ 348
Db 319 TLPPEKQRRKDETYPIKSPTFAGGLFSISKAYFEHIGTYDNQMEIWGENVMSFRVWQ 378
QY 349 CGVLTHPCSHGVHPFKOAPYSRNKALA---NSVRAAEVWMDPEKELYHRN---P 400
Db 379 CGQLIIPCSVVGHVFTKSPHTFPKGTSVIARNQVRLAEVWMDSYKLIIFYRNLOQA 438
QY 401 RARLEPFGDVTERKQLDKLQCKDFKWFLETVYPPELHVPEDRPFGFQMLQNKGLTDYCFD 460
Db 439 MAQESFGDISERLRLREQLRCHNFSWYLNHNVPFVPELDTFTFYGAIKNLG-TNQCILD 497
QY 461 YNPDENQIVGHQVILYLCHGMQONOFFEYTSOKEIRYNTHQPEGCIATVAGMDTILMHL 520
Db 498 VG---ENNRGGKPLIMYCHNLGNQYFEYTSQDLRHNIAK-QLCILHVGSKG-ALGLS 551
QY 521 CEET---APENOKFILOEDGSLFHEQSKKVOAARKSSDSFVPLLRDCTNSD-HQKW 574
Db 552 CHTGKNSQVPKDEEWELAQDLIRNSGGTCLTSQDKK---PAMAPCNPSDPHQLW 605
QY 575 FF 576
Db 606 LF 607
```

## RESULT 13

```
Q8NCL4
ID Q8NCL4 PRELIMINARY; PRT; 622 AA.
AC Q8NCL4;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein FLJ90177.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Mammary gland;
RA Isogai T., Ota T., Nishikawa T., Hayaashi K., Otsuki T., Sugiyama T.,
RA Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Kojima S., Nagahari K.,
RA Masuho Y., Ono T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,
RA Hattori A., Okumura K., Iwayanagi T., Ninomiya K.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK074658; BAC11118.1; -.
DR InterPro; IPR001173; Glyco_trans_2.
DR InterPro; IPR000772; Ricin B lectin.
DR Pfam; PF00535; Glycos transf 2; 1.
DR Pfam; PF00652; Ricin B lectin; 1.
DR SMART; SM00458; RICIN_1.
DR PROSITE; PS50231; RICIN_B_LECTIN; 1.
KW Hypothetical protein.
```

Query Match 37.9%; Score 1185.5; DB 4; Length 622;  
Best Local Similarity 45.9%; Pred. No. 5.3e-88;  
Matches 249; Conservative 91; Mismatches 163; Indels 39; Gaps 14;

```
QY 58 PGRPEVMPRPVPPANALGARGEAVR---LQIQGEELRLQESVRLHQINILSDRISL 113
Db 96 PAELKPFWERPPQDPNAPGADGKAFQSKWTPLETOE---KEGYKXHCNAPASDRISL 152
QY 114 HRL-PERWNPLCCKEKKY-DYDNLPRTSVLIIFNEAWSTLLRTVSVLETSPIILLEEV 171
Db 153 QRSGLGDTTPPECVDQKFRCPPLATTSVLIIFNEAWSTLLRTVSVLHTTFAILLKEI 212
```

```

QY 172 ILVDDSDREHLKERLANELSGLPKVLIRANKREGLVRARLLGASAAAGDVLTFDCHC 231
Db 213 ILVDDASTEHLKEKLEQYVVKQVVRVVRQBERKGLITARLLGASVAAQAEVLTFDAHC 272
QY 232 ECEHGLEPLLORIEHEESAVVCPVIDIDWNTFFYLGNSEGPQI---GGFDWRLVFTWH 288
Db 273 ECFHGLEPLLARIAEDKTVVSPDIVTIDNTFFFAKPVQGRVHSGRGNFDSLTFGWE 332
QY 289 TVPERIRMQSPVDVIRSPMTAGGLFAVSKYFFYLGSYDTGMEVWGGNLEFSFRWQ 348
Db 333 TLPPEHQRKQKDBTYPIKSPFTAGGLFSISKSYFEHIGTYDQMEIWGGENVEMSPRVQ 392
QY 349 CGGVLETHPCSHVGHVPKQAPYSRNKALA-----NSVRAAEVMDSEFKELYHNRPA-- 402
Db 393 CGGLEIIPCSVVGHVFTKSPHTFPKGTSVIARNQVRLAEVWMDSYKKIFYRNLQAOK 452
QY 403 --RLEPRGDVTERKQLRDKLQCKDFKWFLETVVPELHVPEDRPGFQGLMQLNGLTDCFD 460
Db 453 MTQEKSGDISERQLREQHLCHNFPSWYLNHVPFEMFVDPDLTTFYGAIKNLG-TNQCLD 511
QY 461 YNPPDENQIVGHQVILYLCHGMGNQFFEYTSQKEIRYNTHQPEGCIABAGMDTLIMHL 520
Db 512 VG---ENNRGGKPLIMYSCHGLGNQYFEYTTQDRLRHNIAT-RLCLHVSKG--ALGLGS 565
QY 521 CEET-----APENQKFILOEDGSLFHEQSKKCVQAARKESDSFVPLLRDCTNSD-HOKW 574
Db 566 CHFTGKNSQVPKDEWEALAQDLIRNSGSGTCLTSQDKK-----PAMAPCNPSDPHQLW 619
QY 575 FF 576
Db 620 LF 621

RESULT 14
QYU15 PRELIMINARY; PRT; 622 AA.
AC QYU15 PRELIMINARY; PRT; 626 AA.
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE UDP-GalNac:polypeptide N-acetylglactosaminyltransferase.
GN GALNAC-T6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Gastric;
RX MEDLINE=9395099; PubMed=10464263;
RA Bennett E.P., Hassan H., Mandel U., Hollingsworth M.A., Akisawa N.,
RA Ikematsu Y., Merckx G., Geurts van Kessel A., Olofsson S., Clausen H.,
RT "Cloning and Characterization of a Close Homologue of Human UDP-N-
RT acetyl--D-galactosamine:Polypeptide N-Acetylglactosaminyltransferase-
RT T3, Designated GalNac-T6. Evidence for genetic but not functional
RT redundancy."
RL J. Biol. Chem. 274:25362-25370(1999).
DR EMBL; Y08565; CAA69876.1; -.
DR Genbank; HGNC:4128; GALN6.
DR InterPro; IPR001173; Glyco_trans_2.
DR InterPro; IPR000772; Ricin_B_lectin.
DR Pfam; PF00535; Glycos_transf_1.
DR Pfam; PF00652; Ricin_B_lectin; 2.
DR SMART; SM00458; RICIN; 1.
DR PROSITE; PS0231; RICIN_B_LCTIN; 1.
KW Transferase.
SQ SEQUENCE 622 AA; 71168 MW; 3944AFF2AF0916EF CRC64;

Query Match 37.9%; Score 1182.5; DB 4; Length 622;
Best Local Similarity 45.8%; Pred. No. 9.3e-88;
Matches 248; Conservative 91; Mismatches 164; Indels 39; Gaps 14;

```

Search completed: December 4, 2003, 19:43:42  
Job time : 41 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2003, 19:38:37 ; Search time 47 Seconds  
(without alignments)  
1962.131 Million cell updates/sec

Title: US-10-074-527-2

Perfect score: 3124

Sequence: 1 MWGTAARRCPRELRRGREA.....LLRDCTNSDQKWFKKERML 581

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_19Jun03.\*

1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
24: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3124	100.0	581	23	ABG32510 Human glycosyltran
2	3124	100.0	581	23	AAE25019 Human drug metabol
3	1788	57.2	355	22	ABG04767 Novel human diagn
4	1638	52.4	578	24	ABU07527 Human N-acetylglala
5	1369	43.8	256	21	AAE53452 Human colon cancer
6	1309	41.9	315	23	ABE89881 Human polypeptide
7	1309	41.9	316	22	AAE74101 Human colon cancer
8	1304	41.7	290	21	AAE24470 Human secreted pro
9	1187.5	38.0	622	22	AAE79050 Human protein SEQ

10	1187.5	38.0	622	22	AAE38859 Human polypeptide
11	1187.5	38.0	632	22	AAE80034 Human protein SEQ
12	1187.5	38.0	632	22	AAE40645 Human polypeptide
13	1185.5	37.9	622	22	AAE93285 Human polypeptide,
14	1182.5	37.9	622	23	ABJ05568 Breast cancer-asso
15	1142.5	36.6	633	19	AAE34470 Human N-acetylglala
16	1122.5	35.9	556	23	ABG32379 Novel human enzyme
17	1101.5	35.3	561	21	AAE43561 Human cancer assoc
18	1100.5	35.2	559	16	AAE66397 Cattle GalNAC-tran
19	1100.5	35.2	559	16	AAE66401 GalNAC-transferase
20	1100.5	35.2	559	18	AAE16484 Bovine N-acetylgl
21	1100	35.2	517	16	AAE66402 GalNAC-transferase
22	1100	35.2	517	18	AAE16489 Honeybee mellitin
23	1076.5	34.5	630	22	ABE64271 Drosophila melanog
24	1058	33.9	478	24	ABE41518 Human DITP protei
25	1057	33.8	590	22	ABE59763 Drosophila melanog
26	1023	32.7	639	21	AAE99440 Human PRO1564 (UNO
27	1023	32.7	639	22	AAE29205 Human PRO polypept
28	1023	32.7	639	22	ABE73178 UDP-GalNac: polype
29	1023	32.7	639	22	AAE66189 Protein of the inv
30	1023	32.7	639	23	ABE64882 Human protein SEQ
31	1023	32.7	639	23	ABE96444 Human ovarian can
32	1023	32.7	639	24	ABU71293 Human PRO1564 prot
33	1023	32.7	639	24	ABE65750 Human secreted/tra
34	1023	32.7	639	24	ABE66083 Novel human secret
35	1023	32.7	639	24	ABU67587 Human secreted/tra
36	1023	32.7	639	24	ABU65445 Human PRO polypept
37	1023	32.7	639	24	ABU58581 Human PRO polypept
38	1023	32.7	639	24	ABU56117 Human secreted/tra
39	1023	32.7	639	24	ABU57112 Human PRO polypept
40	1017	32.6	639	24	ABU10691 Human secreted/tra
41	1017	32.6	638	21	AAE78352 Human ZAP-3 protei
42	1013.5	32.4	571	22	AAE24054 Human EST encoded
43	1012	32.4	570	16	AAE83053 N-acetylglalactosam
44	991.5	31.7	603	22	AAU07778 Human novel transf
45	991.5	31.7	603	23	AAU99098 Human glycosyl tra

#### ALIGNMENTS

RESULT 1  
ABG32510  
ID ABG32510 standard; Protein; 581 AA.

XX AC ABG32510;

XX DT 15-NOV-2002 (first entry)

XX DE Human glycosyltransferase 33945.

XX DE Human; enzyme; glycosyltransferase; 33945; atherosclerosis;  
XX KW cardiovascular disorder; ischaemia; atherosclerosis; cancer; tumour;  
XX KW congestive heart failure; endothelial cell disorder; psoriasis;  
XX KW diabetic retinopathy; angina; hypertension; atrial fibrillation;  
XX KW valvular disease; cardiomyopathy; haemangioma; pancreatic disorder;  
XX KW cellular proliferative disorder; differentiative disorder; diabetes;  
XX KW autoimmune disease; haematopoietic neoplastic disorder; leukaemia;  
XX KW Hodgkin's disease; chronic myelogenous leukaemia; inflammatory disease;  
XX KW arthritis; multiple sclerosis; viral infection; liver disorder;  
XX KW liver fibrosis; hepatocellular cancer.

XX OS Homo sapiens.

XX PN WO200264815-A2.

XX PD 22-AUG-2002.

XX PF 14-FEB-2002; 2002WO-US05042.

XX PR 15-FEB-2001; 2001US-269202P.

XX PA (MILL-) MILLENNIUM PHARM INC.

XX Olandt PJ, Meyers RE, Galvin KA;  
 XX WPI; 2002-643491/69.  
 DR N-PSDB; ABS52469.  
 XX New 33945 nucleic acid molecule and polypeptide, useful for diagnosing  
 PT and treating disorders involving aberrant or deficient  
 PT glycosyltransferase function or expression such as atherosclerosis or  
 PT endothelial cell disorders -  
 XX  
 XX Claim 8; Page 100-101; 107pp; English.  
 PS  
 XX The invention relates to an isolated 33945 nucleic acid molecule encoding  
 CC a glycosyltransferase polypeptide, its fragment or allelic variant  
 CC and the encoded protein. Also included are host cells, antibodies,  
 CC producing the protein from cell culture, detecting the  
 CC protein/nucleic acid using probes or binding compounds and identifying  
 CC compounds (modulators) which bind the protein. The methods and  
 CC compositions of the present invention are useful for diagnosing and  
 CC treating disorders involving aberrant or deficient glycosyltransferase  
 CC function or expression such as atherosclerosis, cardiovascular  
 CC disorders (e.g. ischaemia, atherosclerosis, congestive heart failure)  
 CC endothelial cell disorders (e.g. tumours, psoriasis, diabetic  
 CC retinopathy) angina, hypertension, atrial fibrillation,  
 CC valvular disease, cardiomyopathy, haemangiomas, cancers, pancreatic  
 CC disorders, cellular proliferative and/or differentiative disorders,  
 CC autoimmune disorders, haematopoietic neoplastic disorders (e.g.  
 CC leukaemia, Hodgkin's disease, chronic myelogenous leukaemia),  
 CC inflammatory diseases (e.g. diabetes, arthritis, multiple sclerosis),  
 CC viral infection (e.g. Hepatitis B, hepatitis C and herpes simplex  
 CC virus) and liver disorders (e.g. liver fibrosis and hepatocellular  
 CC cancer). Many more diseases and disorders are listed in the  
 CC specification. The present sequence represents the glycosyltransferase  
 CC 33945.  
 XX  
 XX Sequence 581 AA;  
 SQ  
 Query Match 100.0%; Score 3124; DB 23; Length 581;  
 Best Local Similarity 100.0%; Pred. No. 8.6e-292;  
 Matches 581; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MWGTTARRRRCPRELRGREALLVLLALLAGLGSVLRQAQAGAGAAEPGPPTRPGR 60  
 DB 1 MWGTTARRRRCPRELRGREALLVLLALLAGLGSVLRQAQAGAGAAEPGPPTRPGR 60  
 QY 61 REPVMPRPVPANALGARGAVALRLOQEEELRLOQEEELRLOQEEELRLOQEEELR 120  
 DB 61 REPVMPRPVPANALGARGAVALRLOQEEELRLOQEEELRLOQEEELRLOQEEELR 120  
 QY 121 WNPCKEKKYDYNLPRTSVIAFYNEAWSTLLRTVYVLETSPIILLEEVILVDDYSDR 180  
 DB 121 WNPCKEKKYDYNLPRTSVIAFYNEAWSTLLRTVYVLETSPIILLEEVILVDDYSDR 180  
 QY 181 EHLKERLANELSGPKVRLIRANKREGILVRARLIGASAARGDVLTFDLCHEGHWLEP 240  
 DB 181 EHLKERLANELSGPKVRLIRANKREGILVRARLIGASAARGDVLTFDLCHEGHWLEP 240  
 QY 241 LLQRIHEESAVVCFVIDVDMNTFEYLGNSGEPQIGFDWRLVFTWHTVPERIRMQS 300  
 DB 241 LLQRIHEESAVVCFVIDVDMNTFEYLGNSGEPQIGFDWRLVFTWHTVPERIRMQS 300  
 QY 301 PVDVIRSPFMAGGLFAVSKKYFEYLGSDYDTGMEVWGGENLEFSPRIWCCGVLETHPCSH 360  
 DB 301 PVDVIRSPFMAGGLFAVSKKYFEYLGSDYDTGMEVWGGENLEFSPRIWCCGVLETHPCSH 360  
 QY 361 VGHVFPKQAPYSRNKALANSVRAAEVWMDPEKELYIYHNPRLARPEPDGVTTERKQLRDKL 420  
 DB 361 VGHVFPKQAPYSRNKALANSVRAAEVWMDPEKELYIYHNPRLARPEPDGVTTERKQLRDKL 420  
 QY 421 QCKDFKWFLETVPYELHVPEDRPGFFGMLQNKGLTDYCFDYNPPDENQIVGHQVILYLCH 480  
 DB 421 QCKDFKWFLETVPYELHVPEDRPGFFGMLQNKGLTDYCFDYNPPDENQIVGHQVILYLCH 480

QY 481 GMGQNOFFEYTSQKEIRYNTHQPEGCIATVEAGMDTILMHLCEBETAPENQKFIQEDGSLF 540  
 DB 481 GMGQNOFFEYTSQKEIRYNTHQPEGCIATVEAGMDTILMHLCEBETAPENQKFIQEDGSLF 540  
 QY 541 HQSKKCVQAARKESDVSFVPLLRDCTNSDQKQWFFKERM 581  
 DB 541 HQSKKCVQAARKESDVSFVPLLRDCTNSDQKQWFFKERM 581

RESULT 2  
 AAE25019  
 ID AAE25019 standard; Protein; 581 AA.  
 XX AC AAE25019;  
 XX 30-OCT-2002 (first entry)  
 DE Human drug metabolising enzyme (DME-4).  
 XX Human; drug metabolising enzyme; autoimmune; inflammatory disorder;  
 KW acquired immunodeficiency syndrome; AIDS; atherosclerosis; psoriasis;  
 KW proliferative disorder; arteriosclerosis; cirrhosis; hepatitis; cancer;  
 KW asthma; neurological disorder; Alzheimer's disease; Huntington's disease;  
 KW dementia; Parkinson's disease; developmental disorder; anaemia; adenoma;  
 KW drug screening; endocrine disorder; conjunctivitis; glaucoma; cataract;  
 KW renal tubular acidosis; eye disorder; epilepsy; thrombosis; peptic ulcer;  
 KW anorexia; metabolic disorder; cystic fibrosis; diabetes; liver disorder;  
 KW goitre; gastrointestinal disorder; gene therapy; virucide; anticoagulant;  
 KW anticonvulsant; nootropic; enzyme; DME-4.  
 XX Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..34  
 FT Protein 35..581  
 FT Peptide 1..39  
 FT Protein 40..581  
 FT Domain 14..41  
 FT /label= Signal\_peptide  
 FT /note= "Mature human DME-4"  
 FT /label= Signal\_peptide  
 FT /note= "Mature human DME-4"  
 FT /note= "Transmembrane domain"  
 XX WO200246426-A2.  
 PN 13-JUN-2002.  
 XX 04-DEC-2001; 2001WO-US47429.  
 XX 08-DEC-2000; 2000US-254308P.  
 PR 15-DEC-2000; 2000US-256189P.  
 PR 21-DEC-2000; 2000US-257713P.  
 PR 19-JAN-2001; 2001US-262706P.  
 PR 02-FEB-2001; 2001US-266020P.  
 XX (INCY-) INCYTE GENOMICS INC.  
 PA Sanjanwala MM, Yao MG, Au-Young J, Baughn MR, Arvizu C, Ring HZ;  
 PI Lee EA, Ding L, Hafalia AJA, Tang YT, Yue H, Tribouley CM;  
 PI Lu DAM, Lal PG, Warren BA, Yang J, Wallia NK, Nguyen DB;  
 PI Gandhi AR, Lu Y, Ison CH;  
 XX WPI; 2002-519668/55.  
 DR N-PSDB; AAD40568.  
 XX Novel human drug metabolizing polypeptide, useful in diagnosis,  
 PT prevention or treatment of autoimmune/inflammatory, cell proliferative,  
 PT neurological, developmental, endocrine, metabolic and gastrointestinal  
 PT disorders -  
 XX Claim 59; Page 146-147; 169pp; English.  
 PS



XX The invention relates to an isolated human drug metabolising enzyme (DME)  
CC and its nucleotide. DME is useful for diagnosing, treating or preventing  
CC disorders associated with aberrant expression of DME, where the disorders  
CC are selected from autoimmune/inflammatory disorder such as acquired  
CC immunodeficiency syndrome (AIDS), asthma, atherosclerosis, psoriasis,  
CC uveitis; a cell proliferative disorder such as arteriosclerosis,  
CC cirrhosis, hepatitis, and cancer; a neurological disorder such as  
CC Alzheimer's disease, Huntington's disease, dementia, Parkinson's disease;  
CC a developmental disorder such as renal tubular acidosis, epilepsy,  
CC anaemia; an endocrine disorder such as adenoma, thrombosis and  
CC infections; an eye disorder such as conjunctivitis, glaucoma, cataract;  
CC metabolic disorder such as cystic fibrosis, diabetes and goitre; a  
CC gastrointestinal disorder such as anorexia, peptic ulcer; and liver  
CC disorders. DME is useful in a number of drug screening techniques and to  
CC analyse the proteome of a tissue or cell type. The invention is useful  
CC for creating knock-in humanised animals or transgenic animals to model  
CC human diseases, in somatic or germline gene therapy, to generate a  
CC transcript image of a tissue or cell type, for detecting differences in  
CC the chromosomal location due to translocation, inversion, etc. among  
CC normal, carrier or affected individuals, and as hybridisation probes for  
CC mapping naturally occurring genomic sequences. The present sequence is  
CC human DME-4.  
XX

SQ Sequence 581 AA;

Query Match 100.0%; Score 3124; DB 23; Length 581;  
Best Local Similarity 100.0%; Pred. No. 8.6e-292;  
Matches 581; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MWGTARRCPRELRGREALLVLLALLAGLGSVLAQRGAGAGAAEPGPPRPRGR 60  
Db 1 MWGTARRCPRELRGREALLVLLALLAGLGSVLAQRGAGAGAAEPGPPRPRGR 60  
Qy 61 REPVMPPPANALGARGAVRLQGEELRLOEESVRLHQINILSDRSLHRRPER 120  
Db 61 REPVMPPPANALGARGAVRLQGEELRLOEESVRLHQINILSDRSLHRRPER 120  
Qy 121 WNPCKEKKDYDNLPRTSVIAFYNEAWSLTLLRTVSVLETSPIILLEEVILVDDYSDR 180  
Db 121 WNPCKEKKDYDNLPRTSVIAFYNEAWSLTLLRTVSVLETSPIILLEEVILVDDYSDR 180  
Qy 181 EHLKERLANELSGLPKVLRLRANKREGILVRALLGASAAAGDVLTFLDCHCEHGWLEP 240  
Db 181 EHLKERLANELSGLPKVLRLRANKREGILVRALLGASAAAGDVLTFLDCHCEHGWLEP 240  
Qy 241 LLQRIHEESAIVCPVIDVDMNTFEYLGNSGEPOIGGFOWRLVFTWHTVPERIRMQS 300  
Db 241 LLQRIHEESAIVCPVIDVDMNTFEYLGNSGEPOIGGFOWRLVFTWHTVPERIRMQS 300  
Qy 301 PVDVIRSPWTAGGLFAVSKKYFEYLGSDYDTGMEVWGGENLEFSPRIWQCGGVLETHPCSH 360  
Db 301 PVDVIRSPWTAGGLFAVSKKYFEYLGSDYDTGMEVWGGENLEFSPRIWQCGGVLETHPCSH 360  
Qy 361 VGHVFPKQAPYSRNKALANSVRAAEVWMDPEFKELYHYHNPRLARLEPGDVTERRKQDRKL 420  
Db 361 VGHVFPKQAPYSRNKALANSVRAAEVWMDPEFKELYHYHNPRLARLEPGDVTERRKQDRKL 420  
Qy 421 QCKDFKWFLETVPYDELHVPEDRPGFGMLQNGKLTDCFDYNPDPENQIVGHQVILYLCH 480  
Db 421 QCKDFKWFLETVPYDELHVPEDRPGFGMLQNGKLTDCFDYNPDPENQIVGHQVILYLCH 480  
Qy 481 GMGNQPFYTSQKEIRYNTHQPGGCIAGVAGMDTLIMHLCEETAPENQKFIQEDGSLF 540  
Db 481 GMGNQPFYTSQKEIRYNTHQPGGCIAGVAGMDTLIMHLCEETAPENQKFIQEDGSLF 540  
Qy 541 HEQSKVCQAARKESSDSFVPLLRDCTNSDHQKWFKKERML 581  
Db 541 HEQSKVCQAARKESSDSFVPLLRDCTNSDHQKWFKKERML 581

RESULT 3  
ABG04767

ID ABG04767 standard; Protein; 355 AA.  
XX  
AC ABG04767;  
DT 13-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #4758.  
XX  
DE Human chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
FN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
XX WPI; 2001-639362/73.  
DR N-PSDB; AAS68954.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX  
PS Claim 20; SEQ ID No 35126; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for creating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG0010-ABG0377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 355 AA;  
Query Match 57.2%; Score 1788; DB 22; Length 355;  
Best Local Similarity 98.5%; Pred. No. 2.5e-163;  
Matches 335; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
Qy 65 MPRPPVPANALGARGAVRLQGEELRLOEESVRLHQINILSDRSLHRRPERWNPL 124  
Db 1 MPRPPVPANALGARGAVRLQGEELRLOEESVRLHQINILSDRSLHRRPERWNPL 60  
Qy 125 CKEKKDYDNLPRTSVIAFYNEAWSLTLLRTVSVLETSPIILLEEVILVDDYSDRHLK 184  
Db 61 CKEKKDYDNLPRTSVIAFYNEAWSLTLLRTVSVLETSPIILLEEVILVDDYSDRHLK 120  
Qy 185 ERLANELSGLPKVLRLRANKREGILVRALLGASAAAGDVLTFLDCHCEHGWLEPLQR 244

Db 121 ERLANELSGLPKVKRLIRANKREGLVRARLLGASARGDVLTFDLDCHCEHGLEPLQR 180  
 Qy 245 IHEESAVVCVIDVIDWNTFEYLGNSEGEQIGGFDHRLVFTWHTVPERERIRMQSDV 304  
 Db 181 IHEESAVVCVIDVIDWNTFEYLGNSEGEQIGGFDHRLVFRWHTVPERERIRMQSDV 240  
 Qy 305 IRSPTMAGGLPAVSKKYFEYLGSDYDTGMEVWGGNLEFSFRIWOCGGVLETHPCSHVGHV 364  
 Db 241 IRSPTMAGGLPAVSKKYFEYLGSDYDTGMEVWGGNLEFSFRIWOCGGVLETHPCSHVGHV 300  
 Qy 365 FPKQAPYSRNKALANSVRAAEVWDEFKELYHHRNPRARL 404  
 Db 301 FPKQAPYSRNKALANSVRAAEVWDEFKELYHHRNPRARL 340

## RESULT 4

ABU07527  
 ID ABU07527 standard; Protein; 578 AA.  
 AC ABU07527;  
 XX  
 DT 13-MAR-2003 (first entry)  
 XX  
 DE Human N-acetylgalactosaminyl transferase T4, GalNac T4.  
 XX  
 KW Human; enzyme; GalNac T4; N-acetylgalactosaminyl transferase T4;  
 KW glycosylation; mucin 1; MUC1; vaccine; antiinflammatory; GalNac-T1;  
 KW GalNac-T2; GalNac-T3.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US6465220-B1.  
 XX  
 PD 15-OCT-2002.  
 XX  
 PF 21-DEC-1998; 98US-0217306.  
 XX  
 PR 21-DEC-1998; 98US-0217306.  
 XX  
 PA (GLYC-) GLYCOZYM APS.  
 XX

XX Hassan FH, Clausen H, Bennett EP, Eisenkraetzer D, Gaetgens J;  
 PI WPT; 2003-147066/14.  
 DR N-PSDB; ABX14962.  
 XX

XX Glycosylating MUC1 acceptor substrate, by glycosylating substrate with  
 PT N-acetylgalactosaminyltransferase T1, GalNac-T2 or GalNac-T3, then with  
 PT human GalNac-T4 to glycosylate specific Ser, Thr residues in substrate  
 PT

XX Claim 2; Column 11-16; 10pp; English.

XX The invention relates to glycosylating a MUC1 (mucin 1) acceptor  
 CC substrate, comprising glycosylating the substrate with enzymatically  
 CC active N-acetylgalactosaminyltransferase (GalNac)-T1, GalNac-T2 or  
 CC GalNac-T3, or with GalNac capable of glycosylating MUC1 glycosylation  
 CC sites that can be glycosylated by GalNac-T1, GalNac-T2 or GalNac-T3, and  
 CC glycosylating the substrate with enzymatically active human GalNac-T4 to  
 CC glycosylate specific Ser, Thr positions in the MUC1 substrate.  
 CC The method is used for glycosylating an MUC1 acceptor substrate. The  
 CC glycosylated substrates are useful in preparation of vaccines and  
 CC antiinflammatory agents. GalNac-T4 exhibits a different substrate  
 CC specificity than previously characterised GalNac transferases. The  
 CC activity of GalNac-T4 is unique and specific to glycosylate specific  
 CC serine and threonine residues in MUC1 tandem repeat.  
 CC The present sequence is the human GalNac T4 protein.

XX Sequence 578 AA;

Query Match 52.4%; Score 1638; DB 24; Length 578;  
 Best Local Similarity 56.8%; Pred. No. 1.5e-148;

Matches 324; Conservative 78; Mismatches 150; Indels 18; Gaps 8;  
 Qy 23 VLLALLALAG-----LGSVLRAQAGAGAAEPPTPRRR-----EPVMPRPVPA 72  
 Db 13 LLAFITVAYIFVELLVSTPHASAGAR-ARELGSRRSLDLQKNTEDLSPLLYKKPPADS 71  
 Qy 73 NALGARGEAVRLQLOGEELRQESVRLHQINTYLSDRISLHRLRPERMNPCKEKKYDY 132  
 Db 72 RALGEGKASKLQNLDELKQEEELIERYAINIYLSDRISLHRIEDKRWYCKSOKFNY 131  
 Qy 133 DNPRTSVIIAFNEAWSTLLRTVYSVLETSPIILLSEVILVDDYSDREHLKERLANELS 192  
 Db 132 RTLPTTSVIIAFNEAWSTLLRTIHSVLETSPIALLKEIILVDDLSDRVLYLKTQLEYIS 191  
 Qy 193 GLPKVRLIRANKREGLVRARLLGASARGDVLTFDLDCHCEHGLEPLQRITHEESAV 252  
 Db 192 NLDVRVRLIRANKREGLVRARLLGATFATGVDLTFLYCHCECNGWLEPLERIGRYETAV 251  
 Qy 253 VCPVIDVIDWNTFEYLGNSEGEQIGGFDHRLVFTWHTVPERERIRMQSPVDVIRSPTMAG 312  
 Db 252 VCPVIDTIDWNTFEYMQIGEPWIGGFDHRLVFTQWHSVPKQERDRIRISDIPRSPTMAG 311  
 Qy 313 GLFAVSKKYFEYLGSDYDTGMEVWGGNLEFSFRIWOCGGVLETHPCSHVGHVFPKQAPYS 372  
 Db 312 GLFAVSKKYFYLGTYDTGMEVWGGNLELSFRWOCGGKLETHPCSHVGHVFPKRAPYA 371  
 Qy 373 RNKALANSVRAAEVWDEFKELYHHRNPRARLPFGDVTERRKOLRDLCKDFKWLFTV 432  
 Db 372 RPNFLQNTARAAEVWDEYKEHFYNNPPARKAYAGDISERKLLRELRCKSFDWYLNKV 431  
 Qy 433 YPELHVPEDRPGFGMLQNKGLTDYCFDYNPPDENQIVGHQVILYLCHGNGQNGOFFEYTS 492  
 Db 432 FPNLHVPEDRPGHGAIRSGISSECLDYNSPDNNP-TGANLSLFGCHGCGNGOFFEYTS 490  
 Qy 493 QKEIRYNTHQPEGIAVEAGMDLIMHLCSE---TAPENQKFIQEDGSLFHEQSKCVQ 549  
 Db 491 NKEIRFNS-VTELCAEYPEQNYVMQNCPCPGFPVPANIWHFKEDGTIFPHPSGLCLS 549  
 Qy 550 AARKESSDSFVPLLRDCTNSD-HQKWFPE 578  
 Db 550 AYTPEGRPDVQ-WRTCDALDKNQIWSFEK 578

## RESULT 5

AAB53452  
 ID AAB53452 standard; Protein; 256 AA.  
 XX

XX AAB53452;

XX AC AC  
 XX DT 09-MAR-2001 (first entry)  
 XX

XX Human colon cancer antigen protein sequence SEQ ID NO:992.

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;  
 KW identification; cytostatic; cardioactive; neuroprotective; vulnary;  
 KW immunomodulatory; muscular; gynaecological; gastrointestinal;  
 KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;  
 KW neural disorder; immune system disorder; muscular disorder;  
 KW reproductive disorder; gastrointestinal disorder; renal disorder;  
 KW infectious disease; cardiovascular disorder.

XX Homo sapiens.

XX WO2000055351-A1.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US05883.

XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

XX

PI Rosen CA, Ruben SM;  
XX WPI; 2000-587534/55.  
DR N-PSDB; AAC98209.  
XX  
PT Colon cancer associated gene sequences, referred to as colon cancer  
PT antigens, useful for the treatment, prevention, and diagnosis of colon  
PT disorders such as colon cancer -  
XX  
PS Claim 11; Page 1569-1570; 2104pp; English.  
XX  
XX AAC97991 to AAC98763 encode the human colon cancer associated proteins,  
CC called human colon cancer antigens, given in AAB53234 to AAB54006. The  
CC human colon cancer antigens can have cytostatic, cardioactive, muscular;  
CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,  
CC vulnary, nephrotropic, antineoplastic and antibacterial activities, and  
CC can be used in gene therapy. The colon cancer antigen polynucleotides,  
CC proteins and antibodies to the proteins are useful for the prevention,  
CC treatment and diagnosis of colon disorders, such as colon cancer. The  
CC polynucleotides may be used in diagnostics and research, such as for  
CC chromosome identification, and as hybridisation probes. The proteins  
CC may also be used to prevent diseases such as neural disorders, immune  
CC system disorders, muscular disorders, reproductive disorders,  
CC gastrointestinal disorders, wounds, renal disorders, infectious  
CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and  
CC AAB54007 represent sequences used in the exemplification of the present  
CC invention.  
XX  
XX Sequence 256 AA;  
XX  
Query Match 43.8%; Score 1369; DB 21; Length 256;  
Best Local Similarity 99.6%; Pred. No. 3.8e-123;  
Matches 247; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 238 LEPQLRIHEEAGVCPVIDVIDWNTFEYLGNSGEPQIGGFDWRLVFTWHTVPERER 297  
Db 6 LEPQLRIHEEAGVCPVIDVIDWNTFEYLGNSGEPQIGGFDWRLVFTWHTVPERER 65  
QY 298 MQSPVDVIRPTMAGGLFAVSKYFEYLGSDYDTGMEVWGGNLEFSPRIWCCGGVLETHP 357  
Db 66 MQSPVDVIRPTMAGGLFAVSKYFEYLGSDYDTGMEVWGGNLEFSPRIWCCGGVLETHP 125  
QY 358 CSHVGHVFPKQAPYSRNKALANSVRAAEVWMDFFKELYHYRNPRARLEPFGDVTERKQLR 417  
Db 126 CSHVGHVFPKQAPYSRNKALANSVRAAEVWMDFFKELYHYRNPRARLEPFGDVTERKQLR 185  
QY 418 DKLQCKDFKWFLETVPYLPVPELHVPEDRPGFGMLQNGLTDCFDYNDENQIVGHQVILY 477  
Db 186 DKLQCKDFKWFLETVPYLPVPELHVPEDRPGFGFGMLQNGLTDCFDYNDENQIVGHQVILY 245  
QY 478 LCHGMGQN 485  
Db 246 LCHGMGQN 253  
RESULT 6  
ID ABB89881  
XX ABB89881 standard; Protein; 315 AA.  
XX  
XX ABB89881;  
XX  
DT 24-MAY-2002 (first entry)  
XX  
XX Human polypeptide SEQ ID NO 2257.  
DE  
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
KW antiallergic; hepatocytic; antidiabetic; antiinflammatory; antiulcer;  
KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; human; secreted protein.  
OS  
XX Homo sapiens.  
XX  
PN WO200190304-A2.  
XX  
XX 29-NOV-2001.  
XX  
XX 18-MAY-2001; 2001WO-US16450.  
XX  
XX 19-MAY-2000; 2000US-205515P.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Birse CE, Rosen CA;  
PI WPI; 2002-122018/16.  
DR N-PSDB; ABL90290.  
XX  
XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and  
PT prevention of neural, immune system, muscular, reproductive, and  
PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative  
PT disorders -  
XX  
XX Claim 11; SEQ ID NO 2257; 2081pp + Sequence Listing; English.  
XX  
XX The invention relates to novel genes (ABL89449-ABL90853) and proteins  
CC (ABB9040-ABB9044) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy. The genes are  
CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC and parasitic infections.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at fip.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 315 AA;  
XX  
Query Match 41.9%; Score 1309; DB 23; Length 315;  
Best Local Similarity 99.6%; Pred. No. 3.2e-117;  
Matches 236; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 345 RIWCCGGVLETHPCSHVGHVFPKQAPYSRNKALANSVRAAEVWMDFFKELYHYRNPRARL 404  
Db 79 QIWCCGGVLETHPCSHVGHVFPKQAPYSRNKALANSVRAAEVWMDFFKELYHYRNPRARL 138  
QY 405 EPEGDYTERKQLRDKLQCKDFKWFLETVPYLPVPELHVPEDRPGFGMLQNGLTDCFDYNDEN 464  
Db 139 EPEGDYTERKQLRDKLQCKDFKWFLETVPYLPVPELHVPEDRPGFGFGMLQNGLTDCFDYNDEN 198  
QY 465 DENQIVGHQVILYCHGMGQNPFYTSQKEIRYNTHQPEGCIAGVAGMDTLIMHLCET 524  
Db 199 DENQIVGHQVILYCHGMGQNPFYTSQKEIRYNTHQPEGCIAGVAGMDTLIMHLCET 258  
QY 525 APENQIFILQEGSLPHEOSKCKVQAARKESDSFVPLLRDCTNSDQKQWFFKERM 581  
Db 259 APENQIFILQEGSLPHEOSKCKVQAARKESDSFVPLLRDCTNSDQKQWFFKERM 315  
RESULT 7  
AAG74101  
ID AAG74101 standard; Protein; 316 AA.  
XX  
XX AAG74101;  
XX  
XX AAG74101;  
DT 03-SEP-2001 (first entry)  
XX  
XX Human colon cancer antigen protein SEQ ID NO:4865.

```

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US26524.
XX
PR 29-SEP-1999; 99US-01571137.
XX
PR 03-NOV-1999; 99US-0163280.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
XX WPI; 2001-235357/24.
XX
XX N-PSDB; AAH33532.
XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers -
XX
XX Claim 11; Page 6646-6648; 9803pp; English.
XX
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
XX cancer-associated nucleic acid molecules (N) and proteins (P), where
XX the proteins are collectively known as colon cancer antigens. The colon
XX cancer antigens have cytostatic activity and can be used in gene
XX therapy and vaccine production. N and P may be used in the prevention,
XX diagnosis and treatment of diseases associated with inappropriate P
XX expression. For example, N and P may be used to treat disorders
XX associated with decreased expression by rectifying mutations or deletions
XX in a patient's genome that affect the activity of P by expressing
XX inactive proteins or to supplement the patients own production of P.
XX Additionally, N may be used to produce the colon cancer-associated Ps
XX by inserting the nucleic acids into a host cell and culturing the cell
XX to express the proteins. N and P can be used in the prevention, diagnosis
XX and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
XX and AAG77789 represent sequences used in the exemplification of the
XX present invention.
XX
XX N.B. Pages 666 to 682 and page 7053 of the sequence listing were
XX missing at time of publication, meaning no sequences are present for
XX SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
XX Sequence 316 AA;
XX
XX Query Match 41.9%; Score 1309; DB 22; Length 316;
XX Best Local Similarity 99.6%; Pred. No. 3.2e-117;
XX Matches 236; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX 345 RIWQCGVLETHPCSHVGHVFPKQAPYSRNKALANSVRAAEVWVWDEFKELYHNRPRARL 404
XX :|||||
XX 80 QIWQCGVLETHPCSHVGHVFPKQAPYSRNKALANSVRAAEVWVWDEFKELYHNRPRARL 139
XX :|||||
XX 405 EPPGDVTERKQLDKLQCKDFKWFLETVPYELHVPEDRPGFGMLQNKGLTDYCFDYNPP 464
XX :|||||
XX 140 EPPGDVTERKQLDKLQCKDFKWFLETVPYELHVPEDRPGFGMLQNKGLTDYCFDYNPP 199
XX :|||||
XX 465 DENQIVGHQVILYLCHGNGQNFPEYTSQKEIRYNTHQPEGCI AVEAGMDTLIMHLCET 524
XX :|||||
XX 200 DENQIVGHQVILYLCHGNGQNFPEYTSQKEIRYNTHQPEGCI AVEAGMDTLIMHLCET 259
XX :|||||
XX 525 APENQKFILODGSILFHEQSKKCVQAAKESDSDVFLRLRDTNDSHOKWFFKERM 581
XX :|||||
XX 260 APENQKFILODGSILFHEQSKKCVQAAKESDSDVFLRLRDTNDSHOKWFFKERM 316
XX :|||||
XX
XX RESULT 8
XX AAB24470
XX ID AAB24470 standard; Protein; 290 AA.

```

```

XX AAB24470;
XX
XX 20-NOV-2000 (first entry)
XX
XX Human secreted protein sequence encoded by gene 34 SEQ ID NO:95.
XX
XX Human; secreted protein; cytostatic; antianaemic; antidiabetic;
XX antiinflammatory; ophthalmological; antirheumatic; antiarthritic;
XX antipsoriatic; antiangiogenic; cardiant; anti-HIV; nootropic;
XX neuroprotective; antimicrobial; antiparkinsonian; cancer;
XX immune system disorder; angiogenesis; hyperproliferative disorder;
XX cardiovascular disorder; apoptosis; neurological disease;
XX infectious disease; wound healing.
XX
XX Homo sapiens.
XX
XX WO200035937-A1.
XX
XX 22-JUN-2000.
XX
XX 16-DEC-1999; 99WO-US29950.
XX
XX 17-DEC-1998; 98US-0112809.
XX
XX 18-DEC-1998; 98US-0113006.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Ebner R, Rosen CA, Endress GA, Soppet DR, Ni J;
XX Duan DR, Moore PA, Shi Y, Lafleur DW, Olsen HS, Florence K;
XX
XX WPI; 2000-431566/37.
XX
XX N-PSDB; AAA78414.
XX
XX Forty seven human nucleic acids encoding secreted proteins, useful in
XX the treatment, prevention and diagnosis of cancers, disorders of the
XX immune system, angiogenesis disorders, neurological diseases and
XX hyperproliferative disorders -
XX
XX Claim 11; Page 514-515; 562pp; English.
XX
XX The polynucleotide sequence given in AAA78381 to AAA78432 encode the
XX human secreted proteins given in AAB24437 to AAB24604. Human secreted
XX proteins have activities based on the tissues and cells the genes are
XX expressed in. Examples of activities include: cytostatic; antianaemic;
XX antidiabetic; antiinflammatory; ophthalmological; antirheumatic;
XX antiarthritic; antipsoriatic; antiangiogenic; cardiant; anti-HIV;
XX neuroprotective; antimicrobial and antiparkinsonian.
XX Human secreted protein polynucleotides, polypeptides, antagonists and/or
XX agonists may be useful in treating, preventing, and/or diagnosing other
XX diseases, disorders, and/or conditions such as: (a) cancers; (b)
XX disorders of the immune system; (c) angiogenesis disorders; (d)
XX hyperproliferative disorders; (e) cardiovascular disorders; (f) diseases
XX associated with increase apoptosis; (g) neurological diseases; and
XX (h) infectious diseases. They are also used to promote wound healing.
XX AAA78372 to AAA78380 and AAB24436 represent sequences used in the
XX exemplification of the present invention.
XX
XX Sequence 290 AA;
XX
XX Query Match 41.7%; Score 1304; DB 21; Length 290;
XX Best Local Similarity 99.2%; Pred. No. 8.5e-117;
XX Matches 235; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
XX
XX 345 RIWQCGVLETHPCSHVGHVFPKQAPYSRNKALANSVRAAEVWVWDEFKELYHNRPRARL 404
XX :|||||
XX 53 QIWQCGVLETHPCSHVGHVFPKQAPYSRNKALANSVRAAEVWVWDEFKELYHNRPRARL 112
XX :|||||
XX 405 EPPGDVTERKQLDKLQCKDFKWFLETVPYELHVPEDRPGFGMLQNKGLTDYCFDYNPP 464
XX :|||||
XX 113 EPPGDVTERKQLDKLQCKDFKWFLETVPYELHVPEDRPGFGMLQNKGLTDYCFDYNPP 172
XX :|||||
XX 465 DENQIVGHQVILYLCHGNGQNFPEYTSQKEIRYNTHQPEGCI AVEAGMDTLIMHLCET 524
XX :|||||
XX
XX

```

Matches	249;	Conservative	91;	Mismatch	163;	Indels	39;	Gaps	14;
QY	58	PGRRPPVMPRPVPV	NALCARGEA	VR---	LOLQGE	ELRLQES	VRHLQIN	IVLS	DRISL 113
Db	96	PAELKPF	WERP	PODP	NAGAD	GAQK	SKWTF	LETQ	---
QY	114	HRRLL	-PER	NPLC	KEKKY	-DYD	NLPRTS	VI	IIFYNEAMSTL
Db	153	QRLSGPD	TRPP	CEVQ	DKFR	CP	PLATTS	VI	IIFHNEAMSTL
QY	172	ILVDDY	SDE	HLKE	RLANEL	SLG	LPK	VR	LIRANKREG
Db	213	ILVDDA	STE	HE	KLKE	QLQ	YV	VR	QBERKGL
QY	232	ECHEG	WLE	PLL	QRI	IHE	ESA	V	CPVID
Db	273	ECFHG	WLE	PLL	ARI	AED	KVV	V	SPDIV
QY	289	TVPER	IR	IM	QSP	DV	IR	SPT	WAGGLFA
Db	333	TLP	PEK	QRR	KD	ET	YPI	K	SPTFAG
QY	349	CGGV	LE	TH	PC	SHV	GF	PK	QAPY
Db	393	CGQ	LE	I	PCS	V	GV	H	FR
QY	401	RA	RP	F	G	D	V	T	R
Db	453	MAQ	E	K	S	F	G	D	I
QY	461	YN	P	D	E	N	Q	I	V
Db	512	VG-	-ENN	R	G	K	P	L	I
QY	521	CBET	----	AP	N	Q	F	I	L
Db	566	CH	T	G	K	S	Q	V	P
QY	575	FF	576						
Db	620	LF	621						
RESULT	10								
AA	AM38859								
ID	AA	AM38859	standard;	Protein;	622	AA.			
XX	AA	AM38859							
AC	AA	AM38859							
XX	AA	AM38859							
DT	22-OCT-2001	(first entry)							
XX	Human	polypeptide	SEQ	ID	NO	2004.			
DE	Human;	neotropic;	immunosuppressant;	cytostatic;	gene	therapy;	cancer;		
XX	peripheral	nervous	system;	neuropathy;	central	nervous	system;	CNS;	
XX	Alzheimer's;	Parkinson's	disease;	Huntington's	disease;	haemostatic;			
KW	amyotrophic	lateral	sclerosis;	Shy-Drager	Syndrome;	chemotactic;			
KW	chemokinetic;	thrombolytic;	drug	screening;	arthritis;	inflammation;			
KW	leukaemia.								
XX	Homo	sapiens.							
OS	WO200153312-A1.								
PN	26-JUL-2001.								
XX	26-DEC-2000;	2000WO-US34263.							
XX	21-JAN-2000;	2000US-0488725.							
PR	25-APR-2000;	2000US-0552317.							
PR	09-JUL-2000;	2000US-0598042.							
PR	19-JUL-2000;	2000US-0620312.							
PR	03-AUG-2000;	2000US-0653450.							



Best Local Similarity 45.9%; Pred. No. 4.7e-105; Matches 249; Conservative 91; Mismatches 163; Indels 39; Gaps 14;	
Qy	58 PGRREPVPVPPANALGARGEAVR---LQLOGEELRLOEESVRLHQINILYLSDRISL 113
Db	106 PAELKPFWRPQDPNAPGADGKAFQSKWTPLETQE---KEEGYKXKCFNAPASDRISL 162
Qy	114 HRRL-PRWNPCKEKKY-DYDNLPRTSVIIAFYNEAWSSTLLRTVYSLTSPDILLEEY 171
Db	163 QRSGLPDRPPECVDQKFRCPPLATTSVIIIFHNEAWSSTLLRTVYSLHTTPAILLKEI 222
Qy	172 ILVDYSDREHLKERLANELSGLPKVRIRANKREGLVRLALGASAAAGDVLTLFLOCHC 231
Db	223 ILVDASTEEHLKEKLSQYVKQLQVVRVROEERKGLITARLLGASVAQAEVLTLFLOAHC 282
Qy	232 ECHGWLPLELQRIHEESAVVCFVIDWNTFEYLGNSGEPQI---GGFDMLRVFTWH 288
Db	283 ECFHGMLEPLARIAEDKTVVSPDIVTIDLNTFEPAKPVQGRVHSGRNFDSLTFGWE 342
Qy	289 TVPRERIRMQSPVDVIRSPMTAGGLFAVSKYFEYLGSDYDTGMEVWNGENLEFSFRWQ 348
Db	343 TLPPEKQRKDETYPIKSPFTAGGLFSISKSYFEHIGTYDQMEIWGGENVENSFRVWQ 402
Qy	349 CGGVLETHPCSHVGHVPKQAPYSRNKALA---NSVRAAEVWMDPEKELYHNRN---P 400
Db	403 CGGLEIIPCSVGVHVFRTKSPHTFPKGTSVIARNQVRLAEVWMDSYKKIFYRNLOAAK 462
Qy	401 RARLEPFGDVTERRKQDRKQCKDFKWFLETVYVPELVHVPEDRPGFFGQMLQKGLTYCFD 460
Db	463 MAQEKSGDISERLQREQLHCHNFNSWYLVHNVPEMFVDPDITPTFYGAIKNLG-TNQCILD 521
Qy	461 YNPDPENQIVGHQVILYLCMGQNGQFFEYTSQKEIRYNTHQPEGCIATVAGMDTLIMHL 520
Db	522 VG---ENNRGGKPLIMYSCHGLGNQYFEYTTORDLRHNTAK-QLCLHVSXG--ALGLGS 575
Qy	521 CEET-----APENQKFILOEDGSLFHEQSKKCVQAAKESDSFVPLLRDCTNSD-HQW 574
Db	576 CHFTGKNSQVPKDBEWELAQDLIRNSGSGTCLTSQDKK-----PAMAPCNPSDPHQLW 629
Qy	575 FF 576
Db	630 LF 631
RESULT 12	
AAM40645	
ID	AAM40645 standard; Protein; 632 AA.
AC	AAM40645;
XX	22-OCT-2001 (first entry)
DT	Human polypeptide SEQ ID NO 5576.
DE	Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX	peripheral nervous system; neuropathy; central nervous system; CNS;
XX	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX	leukaemia.
OS	Homo sapiens.
XX	WO200153312-A1.
XX	26-JUL-2001.
XX	26-DEC-2000; 2000WO-US34263.
XX	21-JAN-2000; 2000US-0488725.
XX	25-APR-2000; 2000US-0552317.
XX	09-JUL-2000; 2000US-0598042.
XX	19-JUL-2000; 2000US-0620312.

PR	03-AUG-2000; 2000US-0653450.
PR	14-SEP-2000; 2000US-0662191.
PR	19-OCT-2000; 2000US-0693036.
PR	29-NOV-2000; 2000US-0727344.
PA	(HYSE-) HYSEQ INC.
XX	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI	Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX	WPI; 2001-442253/47.
DR	N-PSDB; AAI59801.
XX	Novel nucleic acids and polypeptides, useful for treating disorders
PT	such as central nervous system injuries -
XX	Example 2; SEQ ID NO 5576; 10078pp; English.
XX	The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC	the encoded polypeptides (AAM38642-AAM42213) with nontropic,
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful
CC	in gene therapy. A composition containing a polypeptide or polynucleotide
CC	of the invention may be used to treat diseases of the peripheral nervous
CC	system, such as peripheral nervous injuries, peripheral neuropathy and
CC	localised neuropathies and central nervous system diseases, such as
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC	utilization of the activities such as: Immune system suppression,
CC	activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC	assays for receptor activity, arthritis and inflammation, leukaemias and
CC	C.N.S disorders.
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification.
XX	Sequence 632 AA;
SQ	Query Match 38.0%; Score 1187.5; DB 22; Length 632;
	Best Local Similarity 45.9%; Pred. No. 4.7e-105;
	Matches 249; Conservative 91; Mismatches 163; Indels 39; Gaps 14;







Db	108	KPVLDRPPQDSNAPGASGAKFTTNLSVBEQKEKGEAKHCFNAFASDRISLHRDLGPD	167
Qy	120	RWNPCKEKEYD-YDNLPTSVIIAFYNEAWSTLLRTVSVLSTSPDILLEEVLVDDYS	178
Db	168	TRPEECIEQFKRCPPLPFTTSVIIVFHNEAWSTLLRTVHVSYPAILLKEILLVDDAS	227
Qy	179	DREHLKERLANELSGLPKVPFLIRANKREGIARARLLGASAAAGDVLTFELDCHCEHGWL	238
Db	228	VDEYLHDKLDEYVKQFSIVKIVRQRERKGLITARLLGATVATAETLTFELDAHCECFYGL	287
Qy	239	EPLLORIHESAVVCVIDIVIDWNTFEYLGNSGEP-----QIGGFDWRLVFTWHTVPER	293
Db	288	EPLLARIAENYTAVSPDIASIDLNTFEF--NKPSPYGSHNHRGNFDWSLSFCWESLPDH	345
Qy	294	ERIMQSPVDVIRSPTNWAGLFAVSKYFEYLSYDTGMEVWGGENLEFSFRIWQCGVL	353
Db	346	EKQRRKDETYPIKTPTFAGGLFSISKEYFEYIGSYDEEMEIWGGENIEMSFYVWQCGQL	405
Qy	354	ETHPCSHVGHVFPKQAPYSRKA-----LANSVRAAEVWMDPEFKELYVHRNPRA----	405
Db	406	EIMPCSVVGHVFRSKSPHSFPGTQVIARNQVRLAEVWMDPEYKEIFYRRNTDAKIVKQK	465
Qy	406	PFGDVTERRKOLRDKLQCKFKWLFETVYPSELHVPEDRPGFPGMLQNKGLTDYCFDYNPPD	465
Db	466	AFGLSKRFEIKHRLRCNFTWYLNNIYPEVYVFDLNPVISGYIKSVG-QPLCLDVG---	521
Qy	466	ENQIVGHQVILYLCHGMGQNPFEYTSQKEIRYN-----THOPEGCIAVEA-----	515
Db	522	ENNOGKPLIMYTCHGIGGNQYFEYSAQHEIRHNIQKELCHAAQGLVQLKACTYKGHKT	581
Qy	516	LIMHLCEETAPENQKFLQEDGSLFHEQSKCVQAARKESSDSFVPLLRDCTNSDH-QKW	574
Db	582	VW-----TGEQIWEIQDKQLLYNPFLLKVCLSANGEH-----PSLVSCNPSDPLQKW	627
Qy	575	FFKE	578
Db	628	ILSQ	631

Result No.	Score	Query Match	Length	DB ID	Description
1	1746	100.0	2745	24	AA040568 Human drug metabol
2	1746	100.0	2850	24	AB952469 Human cDNA encodin
C 3	1730.8	99.1	2290	22	AA126532 Human breast cancer
C 4	1730.8	99.1	2290	22	AA126538 Human breast cancer
5	1208.4	69.2	1259	21	AA098209 Human colon cancer
6	967.4	55.4	1069	23	AA568954 DNA encoding novel
C 7	821.2	47.0	1352	23	ABV233465 Human prostate exp
C 8	821.2	47.0	1352	23	ABV29323 Human prostate exp

## SUMMARIES

## ALIGNMENTS

XX  
PT 30-OCT-2002 (first entry)

Human drug metabolising enzyme (DME-4) cDNA.

Human, drug metabolising enzyme; autoimmune; inflammatory disorder; acquired immunodeficiency syndrome; AIDS; atherosclerosis; psoriasis; proliferative disorder; arteriosclerosis; cirrhosis; hepatitis; cancer; asthma; neurological disorder; Alzheimer's disease; Huntington's disease; dementia; Parkinson's disease; developmental disorder; anaemia; adenoma; drug screening; endocrine disorder; conjunctivitis; glaucoma; cataract; renal tubular acidosis; eye disorder; epilepsy; thrombosis; peptic ulcer; anaemia; metabolic disorder; cystic fibrosis; diabetes; liver disorder; goitre; gastrointestinal disorder; gene therapy; viricide; anticoagulant; anticonvulsant; noctropic; enzyme; DME-4; gene; ss.

**Homo sapiens.**

Key	Location/Qualifiers
CDS	1..1746
	/*tag= a
	/product= "Human DME-4"
sig_peptide	1..102
	/*tag= b
mat_peptide	103..1743
	/*tag= c

```

FT /product= "Mature human DME-4"
FT 1..117
FT /*tag= d
FT mat_peptide
FT 118..1743
FT /*tag= e
FT /product= "Mature human DME-4"
FT
PX WO200246426-A2.
XX
XX 13-JUN-2002.
XX
XX 04-DEC-2001; 2001WO-US47429.
XX
XX 08-DEC-2000; 2000US-254308P.
XX 15-DEC-2000; 2000US-256189P.
XX 21-DEC-2000; 2000US-257113P.
XX 19-JAN-2001; 2001US-262706P.
XX 02-FEB-2001; 2001US-266020P.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Sanjanwala MM, Yao MG, Au-young J, Baughn MR, Arvizu C, Ring HZ;
XX Lee EA, Ding L, Hafalia AJA, Tang YT, Yue H,
XX Lu DAM, Lal PG, Warren BA, Yang J, Wallia NK,
XX Gandhi AR, Lu Y, Ison CH;
XX
XX WPI; 2002-519668/55.
XX P-FSDB; AAE25019.
XX
XX Novel human drug metabolizing polypeptide, useful in diagnosis,
XX prevention or treatment of autoimmune/inflammatory, cell proliferative,
XX neurological, developmental, endocrine, metabolic and gastrointestinal
XX disorders.
XX
XX Claim 72; Page 161-162; 169pp; English.
XX
XX The invention relates to an isolated human drug metabolising enzyme (DME)
XX and its nucleotide. DME is useful for diagnosing, treating or preventing
XX disorders associated with aberrant expression of DME, where the disorders
XX are selected from autoimmune/inflammatory disorder such as acquired
XX immunodeficiency syndrome (AIDS), asthma, atherosclerosis, psoriasis,
XX uveitis; a cell proliferative disorder such as arteriosclerosis,
XX cirrhosis, hepatitis, and cancer; a neurological disorder such as
XX Alzheimer's disease, Huntington's disease, dementia, Parkinson's disease;
XX a developmental disorder such as renal tubular acidosis, epilepsy,
XX anaemia; an endocrine disorder such as adenoma, thrombosis and
XX infections; an eye disorder such as conjunctivitis, glaucoma, cataract;
XX metabolic disorder such as cystic fibrosis, diabetes and goitre; a
XX gastrointestinal disorder such as anorexia, peptic ulcer; and liver
XX disorders. DME is useful in a number of drug screening techniques and to
XX analyse the proteome of a tissue or cell type. The invention is useful
XX for creating knock-in humanised animals or transgenic animals to model
XX human diseases, in somatic or germline gene therapy, to generate a
XX transcript image of a tissue or cell type, for detecting differences in
XX the chromosomal location due to translocation, inversion, etc. among
XX normal, carrier or affected individuals, and as hybridisation probes for
XX mapping naturally occurring genomic sequences. The present sequence is
XX human DME-4 cDNA.
XX
XX Sequence 2745 BP; 704 A; 612 C; 737 G; 692 T; 0 other;
XX
XX Query Match
XX Best Local Similarity 100.0%; Score 1746; DB 24; Length 2745;
XX Matches 1746; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 ATGTGGGGGCGACGCGCGCGCGCGCTGCCCGGGAACTCGCGCGCGCGCGGAGCG 60
XX 1 ATGTGGGGGCGACGCGCGCGCGCGCTGCCCGGGAACTCGCGCGCGCGCGGAGCG 60
XX
XX 61 CTGTTGTGCTCTCGCGCGCTACTGCGGTGGCGGGCTGGGCTCGGTCGCGGCGGAG 120
XX 61 CTGTTGTGCTCTCGCGCGCTACTGCGGTGGCGGGCTGGGCTCGGTCGCGGCGGAG 120

```

```

121 CGTGGGGCGGGCGCGGGCTGCCAGCCGGGAGACCCCGCGGACACCCCGCGCGCGCGGGCGG 180
121 CGTGGGGCGGGCGCGGGCTGCCAGCCGGGAGACCCCGCGGACACCCCGCGCGCGGGCGG 180
181 CGCGAGCCGGTCATGCGCGGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGAG 240
181 CGCGAGCCGGTCATGCGCGGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGAG 240
241 GCGGTGCGGCTGCAGCTGCAGGCGGAGGAGCTCGCGCTGCAGGAGGAGAGCGTGCAGGCTG 300
241 GCGGTGCGGCTGCAGCTGCAGGCGGAGGAGCTCGCGCTGCAGGAGGAGAGCGTGCAGGCTG 300
301 CACCAGATTAACATCTACCTCAGCGACCGCATCTCACTGACCCCGCTTGCAGGAGCGC 360
301 CACCAGATTAACATCTACCTCAGCGACCGCATCTCACTGACCCCGCTTGCAGGAGCGC 360
361 TGGAAACCGCTGTGCAAGAGAGAAATATGATTATGATTATGATTATGATTATGATTATGATT 420
361 TGGAAACCGCTGTGCAAGAGAGAAATATGATTATGATTATGATTATGATTATGATTATGATT 420
421 ATCATAGCATTTTATATGAGGCTGTGTAACCTCTCTTCGGACAGTTTACAGTGTCTT 480
421 ATCATAGCATTTTATATGAGGCTGTGTAACCTCTCTTCGGACAGTTTACAGTGTCTT 480
481 GAGACATCCCCGGATATCTCTGTAGAGAGTGCATCTTGTAGATGATACAGTGTAGATA 540
481 GAGACATCCCCGGATATCTCTGTAGAGAGTGCATCTTGTAGATGATACAGTGTAGATA 540
541 GAGCACCCTGAAGAGCGCTTGGCCCAATGAGCTTTCGGGACTGCCAAGTGGCGCTGATC 600
541 GAGCACCCTGAAGAGCGCTTGGCCCAATGAGCTTTCGGGACTGCCAAGTGGCGCTGATC 600
601 CGCGCAACAGAGAGAGGCGCTGTGCGAGCCCGCTGTGCGGGGGCTGTGCGGGCGAGG 660
601 CGCGCAACAGAGAGAGGCGCTGTGCGAGCCCGCTGTGCGGGGGCTGTGCGGGCGAGG 660
661 GCGGATGTTCTTGACCTTCTCTGGACTGTCACTGTGAGTGCACAGAGGCTGGTGGAGCGG 720
661 GCGGATGTTCTTGACCTTCTCTGGACTGTCACTGTGAGTGCACAGAGGCTGGTGGAGCGG 720
721 CTGCTGAGAGGATCCATGAGAGGAGTGGCGAGTGGTGTGCGCGGTGATGATGATGATC 780
721 CTGCTGAGAGGATCCATGAGAGGAGTGGCGAGTGGTGTGCGCGGTGATGATGATGATC 780
781 GACTTGAACACCTTCGAATACCTTGGGGAATCTCGCGGAGCCCGAGATCGCGGCTTTCGAC 840
781 GACTTGAACACCTTCGAATACCTTGGGGAATCTCGCGGAGCCCGAGATCGCGGCTTTCGAC 840
841 TGGAGGCTGTGTTTCACTGCGCACACAGTTCTCTGAGAGGAGAGGATACGATGCAATCC 900
841 TGGAGGCTGTGTTTCACTGCGCACACAGTTCTCTGAGAGGAGAGGATACGATGCAATCC 900
901 CCGGTGATGTCATGAGTCTCAAAATGCGTGGTGGGCTGTTTGTGTGAGTAAAGAAA 960
901 CCGGTGATGTCATGAGTCTCAAAATGCGTGGTGGGCTGTTTGTGTGAGTAAAGAAA 960
961 TATTTTGAATATCTGGGGTCTTATGATACAGGATGGAAGTTTGGGGGAGAGAAAACCTC 1020
961 TATTTTGAATATCTGGGGTCTTATGATACAGGATGGAAGTTTGGGGGAGAGAAAACCTC 1020
1021 GAATTTTCTTTAGGATCTGGCAGTGTGGGGTCTTGGAAAACACACCCATGTTCCCAT 1080
1021 GAATTTTCTTTAGGATCTGGCAGTGTGGGGTCTTGGAAAACACACCCATGTTCCCAT 1080
1081 GTTGGCCATGTTTCCCAAGAGAGTCCCTACTCCCGCAACAGGCTCTGGCCCAACAGT 1140
1081 GTTGGCCATGTTTCCCAAGAGAGTCCCTACTCCCGCAACAGGCTCTGGCCCAACAGT 1140
1141 GTTGTGCTGAGTGAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
1141 GTTGTGCTGAGTGAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
1201 CGTGCCCGCTTGGAAACCTTTTGGGGATGTGACAGAGAGGAGAGAGTCCCGGACAAAGCTC 1260

```

```
Db 1201 CGTCCCGCTTGGAACTTTTGGGATGTGACAGAGGAGAGCTCCGGGCAAGCTC 1260
Qy 1261 CAGTGAAGACTTTCAAGTGGTCTTGGAGACTGTGTATCCAGAACTGCATGTCCCTGAG 1320
Db 1261 CAGTGAAGACTTTCAAGTGGTCTTGGAGACTGTGTATCCAGAACTGCATGTCCCTGAG 1320
Qy 1321 GACAGGCTGGCTTCTTCCGGATGCTCCAGAACAAAGACTAAACAGACTACTGCTTTGAC 1380
Db 1321 GACAGGCTGGCTTCTTCCGGATGCTCCAGAACAAAGACTAAACAGACTACTGCTTTGAC 1380
Qy 1381 TATAACCTCCCGATGAACACAGATTGGGACACAGAGTCAATCTGTACCTCTGTGCAT 1440
Db 1381 TATAACCTCCCGATGAACACAGATTGGGACACAGAGTCAATCTGTACCTCTGTGCAT 1440
Qy 1441 GGGATGGCCAGAAATCAGTTTTTCCAGTACACGCTCCAGAAAGAAATACGCTATAACACC 1500
Db 1441 GGGATGGCCAGAAATCAGTTTTTCCAGTACACGCTCCAGAAAGAAATACGCTATAACACC 1500
Qy 1501 CACAGGCTGAGGCTGCATTTGTTGGAGACAGGAATGGATACCTTATCATGCTCTC 1560
Db 1501 CACAGGCTGAGGCTGCATTTGTTGGAGACAGGAATGGATACCTTATCATGCTCTC 1560
Qy 1561 TGGGAAGAACTGCCAGAGATCAGAGTTCTCTTGCAGGAGGATGGATCTTTATTT 1620
Db 1561 TGGGAAGAACTGCCAGAGATCAGAGTTCTCTTGCAGGAGGATGGATCTTTATTT 1620
Qy 1621 CACGAACAGTCCAGAAATGTGTCAGGCTCGAGGAGGAGTGCAGTGCAGTTTCGTT 1680
Db 1621 CACGAACAGTCCAGAAATGTGTCAGGCTCGAGGAGGAGTGCAGTGCAGTTTCGTT 1680
Qy 1681 CCACTTTACAGACTGCACCACTCGATCATCAGAAATGGTTCTTCAAAGAGCGCATG 1740
Db 1681 CCACTTTACAGACTGCACCACTCGATCATCAGAAATGGTTCTTCAAAGAGCGCATG 1740
Qy 1741 TTATGA 1746
Db 1741 TTATGA 1746

RESULT 2
ABS52469
ID ABS52469 standard; cDNA; 2850 BP.
AC ABS52469;
XX
DT 15-NOV-2002 (first entry)
XX
DE Human cDNA encoding glycosyltransferase 33945.
XX
KW Human; ss; gene; glycosyltransferase; 33945; atherosclerosis;
KW cardiovascular disorder; ischaemia; atherosclerosis; cancer; tumour;
KW congestive heart failure; endothelial cell disorder; psoriasis;
KW diabetic retinopathy; angina; hypertension; atrial fibrillation;
KW valvular disease; cardiomyopathy; haemangioma; pancreatic disorder;
KW cellular proliferative disorder; differentiative disorder; diabetes;
KW autoimmune disorder; haematopoietic neoplastic disorder; leukaemia;
KW Hodgkin's disease; chronic myelogenous leukaemia; inflammatory disease;
KW arthritis; multiple sclerosis; viral infection; liver disorder;
KW liver fibrosis; hepatocellular cancer.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 81..1826
FT /*tag=
FT /product= "Glycosyltransferase 33945"
FT /note= "This CDS is specifically claimed in claim 1"
XX
PN WO200264815-A2.
XX
PD 22-AUG-2002.
XX
```

```
PF 14-FEB-2002; 2002WO-US05042.
XX
PR 15-FEB-2001; 2001US-269202P.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Olandt PJ, Meyers RE, Galvin KA;
XX
DR WPI; 2002-643491/69.
DR P-PSDB; ABG32510.
XX
XX New 33945 nucleic acid molecule and polypeptide, useful for diagnosing
PT and treating disorders involving aberrant or deficient
PT glycosyltransferase function or expression such as atherosclerosis or
PT endothelial cell disorders
XX
XX Claim 1; Page 97-100; 107pp; English.
XX
CC The invention relates to an isolated 33945 nucleic acid molecule encoding
CC a glycosyltransferase polypeptide, its fragment or allelic variant
CC and the encoded protein. Also included are host cells, antibodies,
CC producing the protein from cell culture, detecting the
CC protein/nucleic acid using probes or binding compounds and identifying
CC compounds (modulators) which bind the protein. The methods and
CC compositions of the present invention are useful for diagnosing and
CC treating disorders involving aberrant or deficient glycosyltransferase
CC function or expression such as atherosclerosis, cardiovascular
CC disorders (e.g. ischaemia, atherosclerosis, congestive heart failure)
CC endothelial cell disorders (e.g. tumours, psoriasis, diabetic
CC retinopathy) angina, hypertension, atrial fibrillation,
CC valvular disease, cardiomyopathy, haemangiomas, cancers, pancreatic
CC disorders, cellular proliferative and/or differentiative disorders,
CC autoimmune disorders, haematopoietic neoplastic disorders (e.g.
CC leukaemia, Hodgkin's disease, chronic myelogenous leukaemia),
CC inflammatory diseases (e.g. diabetes, arthritis, multiple sclerosis),
CC viral infection (e.g. Hepatitis B, hepatitis C and herpes simplex
CC virus) and liver disorders (e.g. liver fibrosis and hepatocellular
CC cancer). Many more diseases and disorders are listed in the
CC specification. The present sequence encodes the glycosyltransferase
CC 33945.
XX
SQ Sequence 2850 BP; 743 A; 637 C; 761 G; 704 T; 5 other;
```

```
Query Match 100.0%; Score 1746; DB 24; Length 2850;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1746; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGTGGGGCGCACGGCGCGCGCGCTGCCCCGGGAACTGCGCGCGCGCGGAGGCG 60
Db 81 ATGTGGGGCGCACGGCGCGCGCGCTGCCCCGGGAACTGCGCGCGCGCGGAGGCG 140
Qy 61 CTGTTGGTGTCTCTGGCGTACTGGCGGTTGGCGCGGCTGGCTGGTGGCGGCGCAG 120
Db 141 CTGTTGGTGTCTCTGGCGTACTGGCGGTTGGCGCGGCTGGCTGGTGGCGGCGCAG 200
Qy 121 CGTGGGGCGCGGGCGCGGGGTGCGGAGCCGGGACCCCGCGCACCCCGCGCGCGG 180
Db 201 CGTGGGGCGCGGGCGCGGGGTGCGGAGCCGGGACCCCGCGCACCCCGCGCGG 260
Qy 181 CGCGAGCCGGTCTATGCGCGCGCGCGCGGTGCGCGGCGAAACCGCTGGCGGCGCAG 240
Db 261 CGCGAGCCGGTCTATGCGCGCGCGCGCGGTGCGCGGCGAAACCGCTGGCGGCGCAG 320
Qy 241 GCGGTGCGGCTGCAGTGCAGGGCGGAGGAGTGGCGGTGCGAGGAGAGCGTGGCGCTG 300
Db 321 GCGGTGCGGCTGCAGTGCAGGGCGGAGGAGTGGCGGTGCGAGGAGAGCGTGGCGCTG 380
Qy 301 CACGAGATTAAACATCTACCTCAGCGACCGCATCTCACTGACCGCGCGCTGCCCGAGCGC 360
Db 381 CACGAGATTAAACATCTACCTCAGCGACCGCATCTCACTGACCGCGCGCTGCCCGAGCGC 440
Qy 361 TGGAAACCCGCTGTGCAAGAGAGAAATATGATTATGATTAATTTGCCAGGACATCTGTT 420
```

Db 441 TGAACCCGCTGTGCAAGAGAGAAATATGATTAATATTTGCCAGGACATCTGTT 500  
Qy 421 ATCATAGCATTTTATATAGAGCTGTGCACTCTCGGACAGTTTACAGTGTCTTT 480  
Db 501 ATCATAGCATTTTATATAGAGCTGTGCACTCTCGGACAGTTTACAGTGTCTTT 560  
Qy 481 GAGACATCCCCGGGATATCTCTAGAGAGAGTGTATCTTTGATAGTACTACAGTGTAG 540  
Db 561 GAGACATCCCCGGGATATCTCTAGAGAGAGTGTATCTTTGATAGTACTACAGTGTAG 620  
Qy 541 GAGACATGAAAGAGCGCTTGGCCAAATGAGCTTTTCGGAGCTGCCAAAGGTGCGCTGATC 600  
Db 621 GAGACATGAAAGAGCGCTTGGCCAAATGAGCTTTTCGGAGCTGCCAAAGGTGCGCTGATC 680  
Qy 601 CGCCCAACAGAGAGAGCGCTTGGCCAAATGAGCTTTTCGGAGCTGCCAAAGGTGCGCGAGG 660  
Db 681 CGCCCAACAGAGAGAGCGCTTGGCCAAATGAGCTTTTCGGAGCTGCCAAAGGTGCGCGAGG 740  
Qy 661 GCGGATGTTCTGACCTTCTCGGACTGTGCTGAGTGCACGAGAGGTGCTGAGAGCGG 720  
Db 741 GCGGATGTTCTGACCTTCTCGGACTGTGCTGAGTGCACGAGAGGTGCTGAGAGCGG 800  
Qy 721 CTGCTGAGAGAGTCCATGAAGAGAGTCCGCACTGTGCTGAGTGCCTCGGCTGATGATGATC 780  
Db 801 CTGCTGAGAGAGTCCATGAAGAGAGTCCGCACTGTGCTGAGTGCCTCGGCTGATGATGATC 860  
Qy 781 GACTGGAACACTTTCGATATCTCGGGAACCTCGGGGAGCCCGAGATCGGCGTTTCGAC 840  
Db 861 GACTGGAACACTTTCGATATCTCGGGAACCTCGGGGAGCCCGAGATCGGCGTTTCGAC 920  
Qy 841 TGGAGGCTGGTGTTCACGTGACACACAGTTCCTGAGAGGAGAGATACGATGCAATCC 900  
Db 921 TGGAGGCTGGTGTTCACGTGACACACAGTTCCTGAGAGGAGAGATACGATGCAATCC 980  
Qy 901 CCCGTCGATGTCATCAGTCTCCAAACATGCTGCTGGGCTGTTTGTGTGAGTAAGAAA 960  
Db 981 CCCGTCGATGTCATCAGTCTCCAAACATGCTGCTGGGCTGTTTGTGTGAGTAAGAAA 1040  
Qy 961 TATTTTGAATATCTGGGGTCTTATGATACAGGAATGGAAGTTTGGGGAGGAGAAACCTC 1020  
Db 1041 TATTTTGAATATCTGGGGTCTTATGATACAGGAATGGAAGTTTGGGGAGGAGAAACCTC 1100  
Qy 1021 GAAATTTCTTTAGATCTGACGATGCTGGGCTTCTGGAACACACCCATGTTCCCAT 1080  
Db 1101 GAAATTTCTTTAGATCTGACGATGCTGGGCTTCTGGAACACACCCATGTTCCCAT 1160  
Qy 1081 GTTGCCCATGTTTCCCAAGCAAGCTCCCTACTCCCGCAACAAAGGCTCTGGCCAAAGT 1140  
Db 1161 GTTGCCCATGTTTCCCAAGCAAGCTCCCTACTCCCGCAACAAAGGCTCTGGCCAAAGT 1220  
Qy 1141 GTTGTGCGAGTGAATGATGATGAATTTAAGAGCTCTACTACCATCGCAACCCC 1200  
Db 1221 GTTGTGCGAGTGAATGATGATGAATTTAAGAGCTCTACTACCATCGCAACCCC 1280  
Qy 1201 CGTGCCCTTTGGAACCTTTTGGGATGTGACAGAGAGAGAGCTCCGGGACAAAGCTC 1260  
Db 1281 CGTGCCCTTTGGAACCTTTTGGGATGTGACAGAGAGAGAGCTCCGGGACAAAGCTC 1340  
Qy 1261 CAGTGTAAGACTTTCAAGTGTCTTGGAGACTGTGATCCAGAACTGCATGTGCTGAG 1320  
Db 1341 CAGTGTAAGACTTTCAAGTGTCTTGGAGACTGTGATCCAGAACTGCATGTGCTGAG 1400  
Qy 1321 GACAGGCTGCTTCTCGGATGCTCCAGAACAAAGACTTAAAGACTACTGCTTTGAC 1380  
Db 1401 GACAGGCTGCTTCTCGGATGCTCCAGAACAAAGACTTAAAGACTACTGCTTTGAC 1460  
Qy 1381 TATAACCTCCCGATGAAACCAAGATTGTGGGACACAGGCTCATTTCTGACTCTGTGAT 1440  
Db 1461 TATAACCTCCCGATGAAACCAAGATTGTGGGACACAGGCTCATTTCTGACTCTGTGAT 1520  
Qy 1441 GGGATGGGCCAGAAATCAGTTTTTTCGATGATACGTTCCCAAGAAAGAAATACGCTATAACACC 1500  
Db 1521 GGGATGGGCCAGAAATCAGTTTTTTCGATGATACGTTCCCAAGAAAGAAATACGCTATAACACC 1580

Qy 1501 CACCAGCCTGAGGCTGCATTGCTGGAGCAGGAATGATACCCCTTATCATGCATCTC 1560  
Db 1581 CACCAGCCTGAGGCTGCATTGCTGGAGCAGGAATGATACCCCTTATCATGCATCTC 1640  
Qy 1561 TCGGAAGAACTCCCCAGAGAAATCAGAGTTTCATCTTGAGAGGATGATCTTTATTT 1620  
Db 1641 TCGGAAGAACTCCCCAGAGAAATCAGAGTTTCATCTTGAGAGGATGATCTTTATTT 1700  
Qy 1621 CACGAACAGTCCCAAGAAATGCTGCCAGGCTGCCAGGAGAGAGTCCGAGTGACAGTTTCGTT 1680  
Db 1701 CACGAACAGTCCCAAGAAATGCTGCCAGGCTGCCAGGAGAGAGTCCGAGTGACAGTTTCGTT 1760  
Qy 1681 CCACTCTTACGAGACTGCACCAACTCGGATCATCAGAAATGGTTCTTCAAAGAGCGCATG 1740  
Db 1761 CCACTCTTACGAGACTGCACCAACTCGGATCATCAGAAATGGTTCTTCAAAGAGCGCATG 1820  
Qy 1741 TTATGA 1746  
Db 1821 TTATGA 1826

RESULT 3  
AAL26522/c  
ID AAL26522 standard; cDNA; 2290 BP.

XX AAL26522;  
AC AC  
XX  
DT 07-DEC-2001 (first entry)  
DE Human breast cancer expressed polynucleotide 18979.  
KW Human; breast cancer; cell marker; cytostatic; ss.  
OS Homo sapiens.  
PN WO200151628-A2.  
XX  
PD 19-JUL-2001.  
XX  
PF 10-JAN-2001; 2001WO-US00798.  
XX  
PR 14-MAR-2000; 2000US-0176077.  
PR 14-MAR-2000; 2000US-0189167.  
PR 24-MAR-2000; 2000US-0192099.  
PR 29-MAR-2000; 2000US-0193480.  
PR 15-MAY-2000; 2000US-0205230.  
PR 09-JUN-2000; 2000US-0211315.  
PR 25-JUL-2000; 2000US-0220534.

(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

Lillie J, Xu Y, Wang Y, Steinmann K;

WPI; 2001-451856/48.

New peptide useful as a marker for the diagnosis of breast cancer

Claim 1; Page 3534-3535; 3695pp; English.

The invention relates to human breast cancer expressed polynucleotides (AAL07544-AAL26789) and methods of assessing whether a patient is afflicted with breast cancer by examining the correlation between the expression of certain markers and the cancerous state of breast cells. The polynucleotides and encoded polypeptides are potential markers for detecting, diagnosing, monitoring, characterizing treating and potentially preventing breast cancer. The polynucleotides and encoded polypeptides are also useful for isolating compounds with cytostatic activity.

Sequence 2290 BP; 508 A; 676 C; 572 G; 528 T; 6 other;

Query Match 99.1%; Score 1730.8; DB 22; Length 2290;

Best Local Similarity 99.8%; Pred. No. 0; Matches 1743; Conservative 0; Mismatches 2; Indels 1; Gaps 1;			
Qy	1	ATGTGGGGCGACAGGGCGCGCGCTGCGCGGGAACTGCGGGCGCGCGCGGGAGCG	60
Db	2212	ATGTGGTGGCGACAGGGCGCGCGCTGCGCGGGAACTGCGGGCGCGCGGGAGCG	2153
Qy	61	CTGTGTGTCTCTGGGGCTACTGGCGTGGCGCGGGCTGGGCTGGGCTGGGCGCGAG	120
Db	2152	CTGTGTGTCTCTGGGGCTACTGGCGTGGCGCGGGCTGGGCTGGGCTGGGCGCGAG	2093
Qy	121	CGTGGGGCGGGCGCGGGCTGCGGAGCGCGGACCCCGCGCACCCCGCGCGCGGGCGG	180
Db	2092	CGTGGGGCGGGCGGGGGCTGCGGAGCGCGGACCCCGCGCACCCCGCGCGGGCGG	2033
Qy	181	CGGAGCGCGTCAATGCGCGCGCGCGCGTGC CGCGGAAACCGCTGGCGCGCGGGCGAG	240
Db	2032	CGGAGCGCGTCAATGCGCGCGCGCGCGTGC CGCGGAAACCGCTGGCGCGCGGGCGAG	1973
Qy	241	CGGGTGGGGCTGCAGCTGCGAGGGCGAGAGCTGCGGCTGCAGAGGAGAGCGTGGCGTG	300
Db	1972	CGGGTGGGGCTGCAGCTGCGAGGGCGAGAGCTGCGGCTGCAGAGGAGAGCGTGGCGTG	1913
Qy	301	CACAGATTAACTACTACCTCAGCGACCGCATCTCACTGCACCGCGCGCTGCCCGAGCGC	360
Db	1912	CACAGATTAACTACTACCTCAGCGACCGCATCTCACTGCACCGCGCGCTGCCCGAGCGC	1853
Qy	361	TGGAACCCCGTGTGCAAGAGAGAAATATGATTATGATAATTTGCCCGAGCAATCTGTT	420
Db	1852	TGGAACCCCGTGTGCAAGAGAGAAATATGATTATGATAATTTGCCCGAGCAATCTGTT	1793
Qy	421	ATCATAGCATTTTATATGAAGCTGTGTCAACTCTCTTCGGACAGTTTACAGTGTCTTT	480
Db	1792	ATCATAGCATTTTATATGAAGCTGTGTCAACTCTCTTCGGACAGTTTACAGTGTCTTT	1733
Qy	481	GAGACATCCCGGATATCTCTGCTAGAGAGATGATCTTGTAGATGATACAGTGTAGTA	540
Db	1732	GAGACATCCCGGATATCTCTGCTAGAGAGATGATCTTGTAGATGATACAGTGTAGTA	1673
Qy	541	GAGCACCCTGAAGAGCGCTTTGGGCAATGAGCTTTGGGCACTGCCCAAGGTGGCGCTGATC	600
Db	1672	GAGCACCCTGAAGAGCGCTTTGGGCAATGAGCTTTGGGCACTGCCCAAGGTGGCGCTGATC	1613
Qy	601	CGGCGCNAAGAGAGAGCGCTGTGCGAGCGCGCTGTGGGGCGCTGTGGGGCGAG	660
Db	1612	CGGCGCNAAGAGAGAGCGCTGTGCGAGCGCGCTGTGGGGCGCTGTGGGGCGAG	1553
Qy	661	GGCGATGTTCTGACCTTCTGAGTGTCTCACTGTGAGTGCACGAGGGTGGCTGGAGCGG	720
Db	1552	GGCGATGTTCTGACCTTCTGAGTGTCTCACTGTGAGTGCACGAGGGTGGCTGGAGCGG	1493
Qy	721	CTGCTGCAGAGGATCCATGAAGAGGAGTGGCGAGTGGTGGCGGCTGATTGATGTGATC	780
Db	1492	CTGCTGCAGAGGATCCATGAAGAGGAGTGGCGAGTGGTGGCGGCTGATTGATGTGATC	1433
Qy	781	GACTGGAACACCTTCGAATACCTGGGAACTCCGGGAGCCCGAGATCGGCGGTTTCGAC	840
Db	1432	GACTGGAACACCTTCGAATACCTGGGAACTCCGGGAGCCCGAGATCGGCGGTTTCGAC	1373
Qy	841	TGGAGGCTGGTGTTCAGTGGGCAACAGTTCTGAGAGGAGAGGATACGGATGCAATCC	900
Db	1372	TGGAGGCTGGTGTTCAGTGGGCAACAGTTCTGAGAGGAGAGGATACGGATGCAATCC	1313
Qy	901	CCCGTGCATGTCTATAGGTCTCAACAAATGGCTGGTGGGCTGTTGCTGTGAGTAAGAA	960
Db	1312	CCCGTGCATGTCTATAGGTCTCAACAAATGGCTGGTGGGCTGTTGCTGTGAGTAAGAA	1253
Qy	961	TATTTTGAATATCTGGGGCTTTATGATACAGGAATGGAAGTTGGGGAGGAGAAACCTC	1020
Db	1252	TATTTTGAATATCTGGGGCTTTATGATACAGGAATGGAAGTTGGGGAGGAGAAACCTC	1193
Qy	1021	GAATTTTCTTTAGGATCTGGGAGTGTGGGGTCTGGAACACACCCATGTTCCCAT	1080

## RESULT 4

AAL26538/c

ID AAL26538 standard; cDNA; 2290 bp.

XX

AC AAL26538;

XX

DT 07-DEC-2001 (first entry)

XX

DE Human breast cancer expressed polynucleotide 18995.

XX

KW Human; breast cancer; cell marker; cytosolic; ss.

XX

OS Homo sapiens.

XX

PN WO200151628-A2.

XX

XX 19-JUL-2001.

XX

PF 10-JAN-2001; 2001WO-US00798.

XX

PR 14-JAN-2000; 2000US-0176077.

PR 14-MAR-2000; 2000US-0189167.

Db	1192	GAATTTTCTTTAGGATCTGGCAGTGTGGGGTCTTGAAAACACACCCATGTTCCCAT	1133
Qy	1081	GTTGGGCATGTTTTCCTCCCAAGCAAGCTCCCTACTCCCGCAACAAAGGCTCTGGCCACAGT	1140
Db	1132	GTTGGGCATGTTTTCCTCCCAAGCAAGCTCCCTACTCCCGCAACAAAGGCTCTGGCCACAGT	1073
Qy	1141	GTTCTGTGAGCTGAAATGATGGAATGAAATTTAAAGAGCTCTACTACCATCGCAACCCC	1200
Db	1072	GTT-GTGAGCTGAAATGATGGAATGAAATTTAAAGAGCTCTACTACCATCGCAACCCC	1014
Qy	1201	CGTGGCGCTTGAAACCTTTTGGGGATGTACAGAGAGGAGGAGCTCCGGGCAAGCTC	1260
Db	1013	CGTGGCGCTTGAAACCTTTTGGGGATGTACAGAGAGGAGGAGCTCCGGGCAAGCTC	954
Qy	1261	CAGTGTAAAGACTTCAAGTGGTCTTTCGGGATGTGTATCCAGAACTGCAATGTGCCTGAG	1320
Db	953	CAGTGTAAAGACTTCAAGTGGTCTTTCGGGATGTGTATCCAGAACTGCAATGTGCCTGAG	894
Qy	1321	GACAGGCTGGCTTCTTTCGGGATGTCTCCAGAAACAAAGGACTAACAGACTACTGCTTTGAC	1380
Db	893	GACAGGCTGGCTTCTTTCGGGATGTCTCCAGAAACAAAGGACTAACAGACTACTGCTTTGAC	834
Qy	1381	TATAACCTCCCGATGAAAACCAAGATTTGGGACACCAAGTCTGTACCTCTGTCTAT	1440
Db	833	TATAACCTCCCGATGAAAACCAAGATTTGGGACACCAAGTCTGTACCTCTGTCTAT	774
Qy	1441	GGATGGGCGCAGAAATCAGTCTTTTCGAGTACACGTCCCGAGAAAGAAATACGCTATAACACC	1500
Db	773	GGATGGGCGCAGAAATCAGTCTTTTCGAGTACACGTCCCGAGAAAGAAATACGCTATAACACC	714
Qy	1501	CACAGGCTGAGGCTGCAATGCTGTGGAAGCAGGAATGGATACCCCTTATCATGCACTCTC	1560
Db	713	CACAGGCTGAGGCTGCAATGCTGTGGAAGCAGGAATGGATACCCCTTATCATGCACTCTC	654
Qy	1561	TGCGAAGAACTGCCCCAGAGAAATCAGAGTTTCAATCTTCAGAGAGGATGGATCTTTATTT	1620
Db	653	TGCGAAGAACTGCCCCAGAGAAATCAGAGTTTCAATCTTCAGAGAGGATGGATCTTTATTT	594
Qy	1621	CACGAAACAGTCCAAAGAAATGTGTCCAGGCTGCGAGGAGGAGTCCGAGTGCAGTTTCGTT	1680
Db	593	CACGAAACAGTCCAAAGAAATGTGTCCAGGCTGCGAGGAGGAGTCCGAGTGCAGTTTCGTT	534
Qy	1681	CCACTTTACGAGACTGCAACCACTCCGATCATCAGAAATGGTTCTTCAAAGAGCGCATG	1740
Db	533	CCACTTTACGAGACTGCAACCACTCCGATCATCAGAAATGGTTCTTCAAAGAGCGCATG	474
Qy	1741	TTATGA 1746	
Db	473	TTATGA 468	





Db	7	CGTGC	CGGAA	CCGCTGGGCGGGAGCGCGGTGGCGCTGCAGCTGCAGGGCGAGGAGC	66
Qy	272	TGCGGCTGCAGGAGGAGCGTGGCGCTGCACACAGATTAACATCTACTCTCAGCGACCGCA	331		
Db	67	TGCGGCTGCAGGAGGAGCGTGGCGCTGCACAGATTAACATCTACTCTCAGCGACCGCA	126		
Qy	332	TCTCACTGCACCGCGCTCCGAGCGCTGGAAACCGCTGTGCAAGAGAGAAATATG	391		
Db	127	TCTCACTGCACCGCGCTCCGCGCGCTGGAAACCGCTGTGCAAGAGAGAAATATG	186		
Qy	392	ATTATGATAATTGCCCAGGACATCTCTTTATCATAGCATTTTATAATGAAGACCTGGTCAA	451		
Db	187	ATTATGATAATTGCCCAGGACATCTCTTTATCATAGCATTTTATAATGAAGACCTGGTCAA	246		
Qy	452	CTCTCCTTCGGACAGTTTACAGTGTCTCTGAGACATCCCGGATATCCTGCTAGAAGAAG	511		
Db	247	CTCTCCTTCGGACAGTTTACAGTGTCTCTGAGACATCCCGGATATCCTGCTAGAAGAAG	306		
Qy	512	TGATCTTGTAGATGACTACAGTGATAGAGACACTTGAAAGGACGCTTGGCCCAATGAGC	571		
Db	307	TGATCTTGTAGATGACTACAGTGATAGAGACACCTGAAGGAGCGCTTGGCCCAATGAGC	366		
Qy	572	TTTTCGGACATGCCAAAGTGGCCCTGATCCGCGCCAAACAAGAGAGAGGGCTGTGTGCAG	631		
Db	367	TTTTCGGACATGCCAAAGTGGCCCTGATCCGCGCCAAACAAGAGAGAGGGCTGTGTGCAG	426		
Qy	632	CCCGGCTGCTGGGGCGCTCTGCGCGAGGGCGATGTTTGACCTTCTCGACTGTCACT	691		
Db	427	CCCGGCTGCTGGGGCGCTCTGCGCGAGGGCGATGTTTGACCTTCTCGACTGTCACT	486		
Qy	692	GTGAGTGCCACGAAGGGTGGCTGGAGCGCTGCTGCAGAGGATCCATGAAGAGGAGTGG	751		
Db	487	GTGAGTGCCACGAAGGGT-GCTGAGCGCTGCTGCAGAGGATCCATGAAGAGGAGTGG	545		
Qy	752	CAGTGTGTGTCGCGGTGATGTGATGCACTGGAAACACCTTCGAATACTCTGGGGAAC	811		
Db	546	CAGTGTGTGTCGCGGTGATGTGATGCACTGGAAACACCTTCGAATACTCTGGGGAAC	605		
Qy	812	CCGGGAGCCCCAGATCGGCGTTTCCAGCTGGAGGCTGGTGTTCACGTGGCACACAGTTC	871		
Db	606	CCGGGAGCCCCAGATCGGCGTTTCCAGCTGGAGGCTGGTGTTCACGTGGCACACAGTTC	665		
Qy	872	CTGAGAGGAGAGGATACGATCCAAATCCCGCGTGCATGTATCATCAGTCTCCAAATGG	931		
Db	666	CTGAGAGGAGAGGATACGATCCAAATCCCGCGTGCATGTATCATCAGTCTCCAAATGG	725		
Qy	932	CTGTGGGCTGTTGTGTGAGTAAGAAATATTTTGAATATCTGGGCTCTTATGATACAG	991		
Db	726	CTGTGGGCTGTTGTGTGAGTAAGAAATATTTTGAATATCTGGGCTCTTATGATACAG	785		
Qy	992	GAATGGAAGTTTGGGAGGAGAAACCTCGAATTTTCTTTAGGATCTGGCAGTGGGTG	1051		
Db	786	GAATGGAAGTTTGGGAGGAGAAACCTCGAATTTTCTTTAGGATCTGGCAGTGGGTG	845		
Qy	1052	GGGTTCCTGGAACAACCCCATGTTTCCCATGTGGCCAATGTTTTCCTCCCAAGCAAGCTCCCT	1111		
Db	846	GGGTTCCTGGAACAACCCCATGTTTCCCATGTGGCCAATGTTTTCCTCCCAAGCAAGCTCCCT	905		
Qy	1112	ACTCCCGCAACAGGCTCTCGCCAAACAGTGTGTGTGAGCTGAAGTATGATGATGAAT	1171		
Db	906	ACTCCCGCAACAGGCTCTCGCCAAACAGTGTGTGTGAGCTGAAGTATGATGATGAAT	965		
Qy	1172	TTAAAGAGCTCTACTACATTCGAAACCCCGTGCCTGGACCTTTTGGGAGTGTGA	1231		
Db	966	TTAAAGAGCTCTACTACATTCGAAACCCCGTGCCTGGACCTTTTGGGAGTGTGA	1025		
Qy	1232	CAGAGAGGAAGCAGCTCTCGGGAACAAGCTCCAGTGTAAAGACTTCAAGTGTCTTGTGAGA	1291		
Db	1026	CAGAGAGGAAGCAGCTCTCGGGAACAAGCTCCAGTGTAAAGACTTCAAGTGTCTTGTGAGA	1085		
Qy	1292	CTGTGTATCAGAACTGCATGTGCTGAGGACAGGCTGCTCTTCGCGATGCTCAGA	1351		
Db	1086	CTGTGTATCAGAACTGCATGTGCTGAGGACAGGCTGCTCTTCGCGATGCTCAGA	1145		

	Qy	Db	Qy	Db
1352	ACAAAGAGCTAACAGACTACTGCTTTGACCTATAACCTCCCGATGAAAAACAGATTCTGG	1411		
1446	ACAAAGAGCTAACAGACTACTGCTTTGACTTAACTCCCGATGAAAAACAGATTCTGG	1205		
1412	GACACAGGTCATTCTGTACTCTGTCTATGGGATGGGCCAGAAATCA	1457		
1206	GACACAGGTCATTCTGTACTCTGTCTATGGGATGGGCCAGAAATGA	1251		

## RESULT 6

AAS68954  
ID AAS68954 standard; cDNA; 1069 BP.

Db 1021 CTTGG 1025

RESULT 7  
ID ABV23465/c  
XX ABV23465 standard; cDNA; 1352 BP.  
XX AC  
XX ABV23465;  
XX  
DT 16-SEP-2002 (first entry)  
XX  
DE Human prostate expression marker cDNA 23456.  
XX  
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
KW pharmacogenomic marker; gene; ss.  
XX  
XX Homo sapiens.  
XX OS  
XX WO200160860-A2.  
XX PN  
XX  
XX 23-AUG-2001.  
XX PD  
XX  
XX 20-FEB-2001; 2001WO-US05171.  
XX PF  
XX 17-FEB-2000; 2000US-183319P.  
XX PR  
XX 16-MAR-2000; 2000US-189862P.  
XX PR  
XX 23-MAY-2000; 2000US-207454P.  
XX PR  
XX 09-JUN-2000; 2000US-211314P.  
XX PR  
XX 18-JUL-2000; 2000US-219007P.  
XX PR  
XX 13-DEC-2000; 2000US-255281P.  
XX  
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX PA  
XX Schlegel R, Endege WO, Monahan JE;  
XX PI  
XX WPI; 2001-662795/76.  
XX DR  
XX  
XX Novel isolated nucleic acid molecule associated with cancerous state of  
PT prostate cells and correlating with presence of prostate cancer, useful  
PT for detecting presence of prostate cancer, stage of prostate cancer -  
XX  
XX Claim 1; Page 4286-4287; 11750pp; English.  
XX  
XX The invention relates to an isolated nucleic acid molecule (I) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
CC specification or its complement. (I) is useful for:  
CC (a) assessing whether a patient is afflicted with prostate cancer;  
CC (b) monitoring the progression of prostate cancer in a patient;  
CC (c) assessing the efficacy of a test compound to inhibit prostate  
CC cancer in a patient;  
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
CC in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound;  
CC (g) determining whether prostate cancer has metastasized in a patient;  
CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
CC patient;  
CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.  
XX  
XX Sequence 1352 BP; 340 A; 335 C; 299 G; 368 T; 10 other;  
SQ

Query Match 47.0%; Score 821.2; DB 23; Length 1352;  
Best Local Similarity 99.5%; Pred. No. 4e-159;  
Matches 834; Conservative 0; Mismatches 3; Indels 1; Gaps 1;  
1304 TGCATCGGCTCCACAAATGGCTGGTGGCTGTTTGGCTGTGAGTAAGAAATATTGA 968  
969 ATATCTGGGGCTTATGATACAGGAATGGAAGTTGGGAGGAGAAACCTCGAATTTTC 1028  
1244 ATATCTGGGGCTTATGATACAGGAATGGAAGTTGGGAGGAGAAACCTCGAATTTTC 1185

QY 1029 CTTTAGGATCTGGCAGTGTGGTGGGTTCTGGAAACACACCCCATGTTCCCATGTTGGCCA 1088  
Db 1184 CTTTAGGATCTGGCAGTGTGGTGGGTTCTGGAAACACACCCCATGTTCCCATGTTGGCCA 1125  
QY 1089 TGTTCCTCCCAAGCAAGCTCCCTACTCCCGCAACAAGGCTCTGGCAACAGTGTTCGTGC 1148  
Db 1124 TGTTCCTCCCAAGCAAGCTCCCTACTCCCGCAACAAGGCTCTGGCAACAGTGTTCGTGC 1066  
QY 1149 AGCTGAAGTATGGATGGAATTTAAAGAGCTCTACTACCATCGCAACCCCGTGCCTCG 1208  
Db 1065 AGCTGAAGTATGGATGGAATTTAAAGAGCTCTACTACCATCGCAACCCCGTGCCTCG 1006  
QY 1209 CTTTGGAACTTTTGGGATGTGACAGAGAGAGAGCTCCGGGACAAGCTCAGTGTAA 1268  
Db 1005 CTTTGGAACTTTTGGGATGTGACAGAGAGAGAGCTCCGGGACAAGCTCAGTGTAA 946  
QY 1269 AGACTTCAAGTGTCTTGGAGACTGTGTATCCAGAACTCATGTGCTGAGGACAGGCC 1328  
Db 945 AGACTTCAAGTGTCTTGGAGACTGTGTATCCAGAACTCATGTGCTGAGGACAGGCC 886  
QY 1329 TGGCTTCTTCGGGATGCTCCAGAACAAAGGACTTAACAGACTACTGCTTGACTATAACCC 1388  
Db 885 TGGCTTCTTCGGGATGCTCCAGAACAAAGGACTTAACAGACTACTGCTTGACTATAACCC 826  
QY 1389 TCCCGATGAAAACACAGATGTGGGACACAGGTCATTCTGTACCTCTGTGATGGATGGG 1448  
Db 825 TCCCGATGAAAACACAGATGTGGGACACAGGTCATTCTGTACCTCTGTGATGGATGGG 766  
QY 1449 CCAGAATCAGTTCCTCGAGTACACGTCGCCAGAAAGAAATACGCTATTAACACCCAGGCC 1508  
Db 765 CCAGAATCAGTTCCTCGAGTACACGTCGCCAGAAAGAAATACGCTATTAACACCCAGGCC 706  
QY 1509 TGAGGCTGCTGCTGTGGAGCAGGAATGGATACCTTATCATGCTCTCTGCGAAGA 1568  
Db 705 TGAGGCTGCTGCTGTGGAGCAGGAATGGATACCTTATCATGCTCTCTGCGAAGA 646  
QY 1569 AACTGCCCCAGAGAAATCAGAAATTCATCTTTGAGGAGGATGGATCTTTATTCAGCAACA 1628  
Db 645 AACTGCCCCAGAGAAATCAGAAATTCATCTTTGAGGAGGATGGATCTTTATTCAGCAACA 586  
QY 1629 GTCACAGAAATGTCCAGGCTCGAGGAGGAGTGCAGTGCAGATTCGTTCCACTCTT 1688  
Db 585 GTCACAGAAATGTCCAGGCTCGAGGAGGAGTGCAGTGCAGATTCGTTCCACTCTT 526  
QY 1689 ACAGAGCTGCACCAACTCGGATCATCAGAAATGGTTCTTCAAAGAGCGCATGTTATGA 1746  
Db 525 ACAGAGCTGCACCAACTCGGATCATCAGAAATGGTTCTTCAAAGAGCGCATGTTATGA 468

RESULT 8  
ABV29323/c  
ID ABV29323 standard; cDNA; 1352 BP.  
XX  
XX AC ABV29323;  
XX  
XX DT 16-SEP-2002 (first entry)  
XX  
XX DE Human prostate expression marker cDNA 29314.  
XX  
XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
KW pharmacogenomic marker; gene; ss.  
XX  
XX OS Homo sapiens.  
XX  
XX PN WO200160860-A2.  
XX  
XX PD 23-AUG-2001.  
XX  
XX 20-FEB-2001; 2001WO-US05171.  
XX  
XX PF 17-FEB-2000; 2000US-183319P.  
XX PR 16-MAR-2000; 2000US-189862P.  
XX PR 25-MAY-2000; 2000US-207454P.

PR 09-JUN-2000; 2000US-211314P.  
PR 18-JUL-2000; 2000US-219007P.  
PR 13-DEC-2000; 2000US-255281P.  
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX Schlegel R, Endege WO, Monahan JE;  
XX WPI; 2001-662795/76.  
XX Novel isolated nucleic acid molecule associated with cancerous state of  
PT prostate cells and correlating with presence of prostate cancer, useful  
PT for detecting presence of prostate cancer, stage of prostate cancer -  
XX Claim 1; Page 6275; 11750pp; English.  
XX The invention relates to an isolated nucleic acid molecule (I) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
CC specification or its complement. (I) is useful for:  
CC (a) assessing whether a patient is afflicted with prostate cancer;  
CC (b) monitoring the progression of prostate cancer in a patient;  
CC (c) assessing the efficacy of a test compound to inhibit prostate  
CC cancer in a patient;  
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
CC in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound;  
CC (g) determining whether prostate cancer has metastasized in a patient;  
CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
CC patient;  
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.  
XX Sequence 1352 BP; 340 A; 335 C; 299 G; 368 T; 10 other;  
SQ  
Query Match 47.0%; Score 821.2; DB 23; Length 1352;  
Best Local Similarity 99.5%; Pred. No. 4e-159;  
Matches 834; Conservative 0; Mismatches 3; Indels 1; Gaps 1;  
QY 909 TGCTATCAGTCTCCCAACAAATGGCTGGTGGCTGTTCTGTGATGAAGAAATATTGA 968  
DB 1304 TGCCATGTGGTCTCCCAACAAATGGCTGGTGGCTGTTCTGTGATGAAGAAATATTGA 1245  
QY 969 ATATCTGGGTCTTATGATACAGGAATGGAAGTTTGGGGAGGAGAAACCTCGAATTTTC 1028  
DB 1244 ATATCTGGGTCTTATGATACAGGAATGGAAGTTTGGGGAGGAGAAACCTCGAATTTTC 1185  
QY 1029 CTTTAGATCTGGCAGTGTGGTGGTCTTGAAACACACACCATGTTCCCATGTGGCCA 1088  
DB 1184 CTTTAGATCTGGCAGTGTGGTGGTCTTGAAACACACACCATGTTCCCATGTGGCCA 1125  
QY 1089 TGTTTCCCAAGCAAGCTCCCTACTCCGCAACAGGCTCTGGCCACAGTGTTCGTGC 1148  
DB 1124 TGTTTCCCAAGCAAGCTCCCTACTCCGCAACAGGCTCTGGCCACAGTGTTCGTGC 1066  
QY 1149 AGCTGAAGTATGGATGGATGAATTTAAAGAGCTCTACTACCATCGCAACCCCGTGGCCG 1208  
DB 1065 AGCTGAAGTATGGATGGATGAATTTAAAGAGCTCTACTACCATCGCAACCCCGTGGCCG 1006  
QY 1209 CTTGGAACCTTTTGGGGATGTGACAGAGAGGAAGCAGCTCCGGGACAGCTCCAGTGTAA 1268  
DB 1005 CTTGGAACCTTTTGGGGATGTGACAGAGAGGAAGCAGCTCCGGGACAGCTCCAGTGTAA 946  
QY 1269 AGACTTCAAGTGGTCTTGGAGACTGTGTATCCAGAACTGATGCTGGCTGAGGACAGGCC 1328  
DB 945 AGACTTCAAGTGGTCTTGGAGACTGTGTATCCAGAACTGATGCTGGCTGAGGACAGGCC 886  
QY 1329 TGGCTTCTTGGGATGCTCCGAAACAAAGGACTTAACAGACTACTCTTGTACTATAACCC 1388  
DB 885 TGGCTTCTTGGGATGCTCCGAAACAAAGGACTTAACAGACTACTCTTGTACTATAACCC 826  
QY 1389 TCCCGATCAAAACCAAGATTTGGGACACAGGTCAATCTGTACTCTCTCATGGATGGG 1448  
DB 825 TCCCGATCAAAACCAAGATTTGGGACACAGGTCAATCTGTACTCTCTCATGGATGGG 766

1449 CCAGAAATCAGTTTTTTCGAGTACACGTCCAGAAAGAAATACGCTATAACACCCAGCC 1508  
DB 765 CCAGAAATCAGTTTTTTCGAGTACACGTCCAGAAAGAAATACGCTATAACACCCAGCC 706  
QY 1509 TGAGGGCTGCATTGCTGTGGAAGCAAGAAATGGATACCCCTTATCATGATCTCTGCGAGA 1568  
DB 705 TGAGGGCTGCATTGCTGTGGAAGCAAGAAATGGATACCCCTTATCATGATCTCTGCGAGA 646  
QY 1569 AACTGCCCCAGAGAAATCAGAAATGTTCTTGCAGGAGGATGGATCTTTATTTCCAGAA 1628  
DB 645 AACTGCCCCAGAGAAATCAGAAATGTTCTTGCAGGAGGATGGATCTTTATTTCCAGAA 586  
QY 1629 GTCCAAAGAAATGTGTCAGGCTCGAGGAGGAGTGGATGACAGATTTTCCTTCCACTCTT 1688  
DB 585 GTCCAAAGAAATGTGTCAGGCTCGAGGAGGAGTGGATGACAGATTTTCCTTCCACTCTT 526  
QY 1689 ACAGAGACTGCACCAACTCGGATCATCAGAAATGGTCTTCAAGAGCGCATGTATGA 1746  
DB 525 ACAGAGACTGCACCAACTCGGATCATCAGAAATGGTCTTCAAGAGCGCATGTATGA 468

RESULT 9  
AAH33532  
ID AAH33532 standard; cDNA; 1517 BP.  
XX AC AAH33532;  
XX 03-SEP-2001 (first entry)  
DE Human colon cancer antigen encoding cDNA SEQ ID NO:588.  
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;  
KW Colorectal carcinoma; ss.  
XX Homo sapiens.  
OS WO200122920-A2.  
PN 05-APR-2001.  
XX 28-SEP-2000; 2000WO-US26524.  
XX 29-SEP-1999; 99US-0157137.  
PR 03-NOV-1999; 99US-0163280.  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA Ruben SM, Barash SC, Birse CE, Rosen CA;  
PI WPI; 2001-235357/24.  
XX P-PSDB; AAG74101.  
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
XX useful for preventing, diagnosing and/or treating colorectal cancers -  
PS Claim 1; Page 2640; 9803pp; English.  
XX AAH32943 to AAH37195 and AAG77788 represent human colon  
CC cancer-associated nucleic acid molecules (N) and proteins (P), where  
CC the proteins are collectively known as colon cancer antigens. The colon  
CC cancer antigens have cytostatic activity and can be used in gene  
CC therapy and vaccine production. N and P may be used in the prevention,  
CC diagnosis and treatment of diseases associated with inappropriate P  
CC expression. For example, N and P may be used to treat disorders  
CC associated with decreased expression by rectifying mutations or deletions  
CC in a patient's genome that affect the activity of P by expressing  
CC inactive proteins or to supplement the patient's own production of P.  
CC Additionally, N may be used to produce the colon cancer-associated Ps  
CC by inserting the nucleic acids into a host cell and culturing the cell  
CC to express the proteins. N and P can be used in the prevention, diagnosis  
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204  
CC and AAB77789 represent sequences used in the exemplification of the



Db 779 CTTCCGGATGCTCCAGAACAAAGGACTAACAGACTACTGCTTTGACTATAACCTCCCGA 838  
Qy 1395 TGAACACGAGTGTGGACACACAGGTCATTCTCTACTCTGTCTATGGGATGGGCCAGAA 1454  
Db 839 TGAACACGAGTGTGGACACACAGGTCATTCTCTACTCTGTCTATGGGATGGGCCAGAA 898  
Qy 1455 TCAGTTTTTCAGTACACGCTCCAGAAAGAAATACGCTATACACCCACACAGCCTGAGGG 1514  
Db 899 TCAGTTTTTCAGTACACGCTCCAGAAAGAAATACGCTATACACCCACACAGCCTGAGGG 958  
Qy 1515 CTGCATTGCTGTGAACGAGGAATGGATACCTTTATCATGCATCTCTGCGAAGAACTGC 1574  
Db 959 CTGCATTGCTGTGAACGAGGAATGGATACCTTTATCATGCATCTCTGCGAAGAACTGC 1018  
Qy 1575 CCCAGAGATCAGAAAGTTCATCTTCAGGAGGATGGATCTTTATTTACAGACAGTCCAA 1634  
Db 1019 CCCAGAGATCAGAAAGTTCATCTTCAGGAGGATGGATCTTTATTTACAGAACTCCAA 1078  
Qy 1635 GAAATGTGTCCAGGCTGCGAGGAAGGAGTCCAGTGACAGTTTCTGTTCCACTCTTTACGAGA 1694  
Db 1079 GAAATGTGTCCAGGCTGCGAGGAAGGAGTCCAGTGACAGTTTCTGTTCCACTCTTTACGAGA 1138  
Qy 1695 CTGCACCAACTCGGATCATCAGAAATGGTCTTCTCAAGAGCGCATGTTATGA 1746  
Db 1139 CTGCACCAACTCGGATCATCAGAAATGGTCTTCTCAAGAGCGCATGTTATGA 1190

RESULT 11

AAV78414  
ID AAA78414 standard; cDNA; 1517 BP.  
XX AC AAA78414;  
XX DT 20-NOV-2000 (first entry)  
XX DE Human secreted protein gene 34 SEQ ID NO:44.  
XX KW Human; secreted protein; cytostatic; antianaemic; antidiabetic;  
XX KW antiinflammatory; ophthalmological; antirheumatic; antiarthritic;  
XX KW antiproliferative; angiogenic; cardiant; anti-HIV; neutrotic;  
XX KW neutrotic; antimicrobial; antiparkinsonian; cancer;  
XX KW immune system disorder; angiogenesis; hyperproliferative disorder;  
XX KW cardiovascular disorder; apoptosis; neurological disease;  
XX KW infectious disease; wound healing; ss.  
XX OS Homo sapiens.  
XX PN WO200035937-A1.  
XX PD 22-JUN-2000.  
XX PF 16-DEC-1999; 99WO-US29950.  
XX PR 17-DEC-1998; 98US-0112809.  
XX PR 18-DEC-1998; 98US-0113006.  
XX XX  
XX PA (HUMA-) HUMAN GENOME SCI INC.  
XX PI Ruben SM, Ebner R, Rosen CA, Endress GA, Soppet DR, Ni J;  
XX PI Duan DR, Moore PA, Shi Y, Lafleur DW, Olsen HS, Florence K;  
XX XX  
XX DR WPI; 2000-431566/37.  
XX DR P-PSDB; AAB24470.  
XX XX  
XX PT Forty seven human nucleic acids encoding secreted proteins, useful in  
XX PT the treatment, prevention and diagnosis of cancers, disorders of the  
XX PT immune system, angiogenesis disorders, neurological diseases and  
XX PT hyperproliferative disorders -  
XX PS Claim 1; Page 465-466; 562pp; English.  
XX CC The polynucleotide sequence given in AAA78381 to AAA78432 encode the

CC human secreted proteins given in AAB24437 to AAB24604. Human secreted  
CC proteins have activities based on the tissues and cells the genes are  
CC expressed in. Examples of activities include: cytostatic; antianaemic;  
CC antidiabetic; antiinflammatory; ophthalmological; antirheumatic;  
CC antiarthritic; antiproliferative; angiogenic; cardiant; anti-HIV;  
CC neutrotic; neutrotic; antimicrobial; antiparkinsonian;  
CC human secreted protein polynucleotides, polypeptides, antagonists and/or  
CC agonists may be useful in treating, preventing, and/or diagnosing other  
CC diseases, disorders, and/or conditions such as: (a) cancers; (b)  
CC disorders of the immune system; (c) angiogenesis disorders; (d)  
CC hyperproliferative disorders; (e) cardiovascular disorders; (f) diseases  
CC associated with increase apoptosis; (g) neurological diseases; and  
CC (h) infectious diseases. They are also used to promote wound healing.  
CC AAA78372 to AAA78380 and AAB24436 represent sequences used in the  
CC exemplification of the present invention.

XX Sequence 1517 BP; 366 A; 360 C; 363 G; 424 T; 4 other;

Query Match 40.8%; Score 711.6; DB 21; Length 1517;  
Best Local Similarity 99.9%; Pred. No. 1.3e-136;  
Matches 711; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Qy 1035 GATCTGGCAGTGTGGGGTCTTGGAAACACACCCATGTTCCCATGTTGGCCATGTTT 1094  
Db 479 GATCTGGCAGTGTGGGGTCTTGGAAACACACCCATGTTCCCATGTTGGCCATGTTT 538  
Qy 1095 CCCAAGCAGCTCCCTACTCCCGCAACAGGCTCTGGCCAAACAGTCTTGTGAGCTGA 1154  
Db 539 CCCAAGCAGCTCCCTACTCCCGCAACAGGCTCTGGCCAAACAGTCTTGTGAGCTGA 598  
Qy 1155 AGTATGGATGGATGAATTTAAAGAGCTCTACTACCATCGCAACCCCGTCCCGTGGGA 1214  
Db 599 AGTATGGATGGATGAATTTAAAGAGCTCTACTACCATCGCAACCCCGTCCCGTGGGA 658  
Qy 1215 ACCTTTGGGATGTGACAGAGGAGAGCTCCGGCAACAGTCCAGTCTGTAAGACTT 1274  
Db 659 ACCTTTGGGATGTGACAGAGGAGAGCTCCGGCAACAGTCCAGTCTGTAAGACTT 718  
Qy 1275 CAAGTGTCTTCGAGACTGTGTATCCAGAACTGCTGCTGCTGAGGACAGGCTGCTT 1334  
Db 719 CAAGTGTCTTCGAGACTGTGTATCCAGAACTGCTGCTGCTGAGGACAGGCTGCTT 778  
Qy 1335 CTTTCGGGATGCTCCAGAAACAAAGGACTTAACAGACTACTGCTTTGACTATAACCTCCGA 1394  
Db 779 CTTTCGGGATGCTCCAGAAACAAAGGACTTAACAGACTACTGCTTTGACTATAACCTCCGA 838  
Qy 1395 TGAACACGAGTGTGGGACACAGGTCATTCTGTACTCTGTCTGATGGATGGGCCAGAA 1454  
Db 839 TGAACACGAGTGTGGGACACAGGTCATTCTGTACTCTGTCTGATGGATGGGCCAGAA 898  
Qy 1455 TCAGTTTTTCAGTACACGCTCCAGAAAGAAATACGCTATAACACCCACAGCCTGAGGG 1514  
Db 899 TCAGTTTTTCAGTACACGCTCCAGAAAGAAATACGCTATAACACCCACAGCCTGAGGG 958  
Qy 1515 CTGCATTGCTGTGAACGAGGAATGGATACCTTTATCATGCATCTCTGCGAAGAACTGC 1574  
Db 959 CTGCATTGCTGTGAACGAGGAATGGATACCTTTATCATGCATCTCTGCGAAGAACTGC 1018  
Qy 1575 CCCAGAGATCAGAAAGTTCATCTTCAGGAGGATGGATCTTTATTTACAGACAGTCCAA 1634  
Db 1019 CCCAGAGATCAGAAAGTTCATCTTCAGGAGGATGGATCTTTATTTACAGACAGTCCAA 1078  
Qy 1635 GAAATGTGTCCAGGCTGCGAGGAAGGAGTCCAGTGACAGTTTCTGTTCCACTCTTTACGAGA 1694  
Db 1079 GAAATGTGTCCAGGCTGCGAGGAAGGAGTCCAGTGACAGTTTCTGTTCCACTCTTTACGAGA 1138  
Qy 1695 CTGCACCAACTCGGATCATCAGAAATGGTCTTCTCAAGAGCGCATGTTATGA 1746  
Db 1139 CTGCACCAACTCGGATCATCAGAAATGGTCTTCTCAAGAGCGCATGTTATGA 1190

RESULT 12

ABX14962

ID	ABX14962	standard; DNA; 1737 BP.
XX	AC	
XX	AC	ABX14962;
XX	XX	
XX	DT	13-MAR-2003 (first entry)
XX	XX	
DE	Human	N-acetylgalactosaminyl transferase T4, GalNac T4, gene.
XX	XX	
KW	Human; ds	gene; GalNac T4; N-acetylgalactosaminyl transferase T4;
KW	glycosylation;	mucin 1; MUC1; vaccine; antiinflammatory; GalNac-T1;
KW	GalNac-T2;	GalNac-T3.
XX	XX	
OS	Homo sapiens.	
XX	XX	
XX	Key	Location/Qualifiers
XX	CDS	1..1737
FT	FT	/*tag= a
FT	FT	/product= "GalNac T4"
XX	XX	
PN	US6465220-B1.	
XX	XX	
XX	PD	15-OCT-2002.
XX	XX	
PP	21-DEC-1998;	98US-0217306.
XX	XX	
PR	21-DEC-1998;	98US-0217306.
XX	XX	
PA	(GLYC-)	GLYCOZYM APS.
XX	XX	
PI	Hassan FH, Clausen H, Bennett BP, Eisenkraetzer D, Gaetgens J;	
XX	XX	
DR	WPI; 2003-147066/14.	
XX	P-PSDB; ABU07527.	
XX	XX	
PT	Glycosylating MUC1	acceptor substrate, by glycosylating substrate with
PT	N-acetylgalactosaminyltransferase T1, GalNac-T2 or GalNac-T3, then with	human GalNac-T4 to glycosylate specific Ser, Thr residues in substrate
PT	XX	
XX	XX	
PS	Example 1; Column 9-12; 10pp; English.	
XX	XX	
CC	The invention relates to glycosylating a MUC1 (mucin 1) acceptor	
CC	substrate, comprising glycosylating the substrate with enzymatically	
CC	active N-acetylgalactosaminyltransferase (GalNac)-T1, GalNac-T2 or	
CC	GalNac-T3, or with GalNac capable of glycosylating MUC1 glycosylation	
CC	sites that can be glycosylated by GalNac-T1, GalNac-T2 or GalNac-T3, and	
CC	glycosylating the substrate with enzymatically active human GalNac-T4 to	
CC	glycosylate specific Ser, Thr positions in the MUC1 substrate.	
CC	The method is used for glycosylating an MUC1 acceptor substrate. The	
CC	glycosylated substrates are useful in preparation of vaccines and	
CC	antiinflammatory agents. GalNac-T4 exhibits a different substrate	
CC	specificity than previously characterised GalNac transferases. The	
CC	activity of GalNac-T4 is unique and specific to glycosylate specific	
CC	serine and threonine residues in MUC1 tandem repeat.	
XX	The present sequence is the human GalNac T4 gene.	
XX	XX	
SQ	Sequence 1737 BP; 482 A; 354 C; 437 G; 464 T; 0 other;	
Query Match 28.4%; Score 496.2; DB 25; Length 1737;		
Best Local Similarity 62.3%; Pred. No. 2e-92;		
Matches 796; Conservative 0; Mismatches 478; Indels 3; Gaps 1		
QY	220	GCCTGGCGCGCGCGCGCGCTGCGGTGCGCTGCAGCTGCAGGGCGAGAGCTCGGCTG 279
Db	217	GCATCTGGGGAGTGGGGGAAGCAGCAAACTCCAGCTCAACGAGGATGAACGTGAAGCAG 276
QY	280	CAGGAGGAGAGCGTGCAGCTGCACACAGATTAACTCTACCTCAGCGACCGCATCTCACTG 339
Db	277	CAAGAAGAACTCATTTGAGAGATACCCATCAATATTACTCAGTGACAGGATTTCCCTG 336
QY	340	CACCGCGCTGCCCGAGCGCTGGAAACCCGCTGTGCAAAGAGAAGAAATATGATTATGAT 399
Db	337	CATCGACACATAGAGGATAAAGAATGTATGAGTGTAAAGTCCCAAGATTCACTATAGG 396



```

QY 1480 AAGAAATACGCTATAA 1496
DB 1474 AAGAAATAGGTTTAA 1490

RESULT 13
ABL81812
XX ABL81812 standard; cDNA; 496 BP.
AC ABL81812;
XX
DT 17-MAY-2002 (first entry)
XX
DE Human ovarian cancer related cDNA clone SEQ ID NO:4790.
XX
KW Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200192581-A2.
XX
PD 06-DEC-2001.
XX
PF 29-MAY-2001; 2001WO-US17756.
XX
PR 26-MAY-2000; 2000US-207484P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Algate PA, Harlocker SL, Jones R;
XX
WPI; 2002-122075/16.
XX
Composition for therapy and diagnosis of ovarian cancer comprising
PT polypeptide of a ovarian tumor polypeptide, polynucleotide encoding
PT polypeptide, antibody specific to polypeptide or T cell expressing
PT polypeptide -
PS Claim 1; SEQ ID 4790; 489pp; English.
XX
The present invention describes a composition (I) comprising: carriers
CC and immunostimulants; and a polypeptide (II) of a ovarian tumour
CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence
CC (S1) from the 10912 nucleotide sequences as given in ABL77023 to
CC ABL87934, (III) encoding (II) having a sequence (S2), a T cell
CC population of (II), or antigen presenting cells that express (II).
CC (I) has cytostatic activity. An oligonucleotide (IV) that hybridises to
CC (S1) can be used for detecting ovarian cancer in a patient's biological
CC sample preferably serum or ovarian tissue. The method comprises
CC contacting a biological sample from a patient with (IV), detecting the
CC amount of polynucleotide hybridising to (IV) and comparing the amount to
CC a predetermined cutoff value and thereby detecting ovarian cancer in the
CC patient, where the amount of polynucleotide hybridising to (IV) is
CC detected preferably by polymerase chain reaction (PCR). (I) comprising
CC (III) and/or (II) is useful for stimulating and/or expanding T cells
CC specific for an ovarian tumour protein comprising contacting T cells
CC with (III) or (II). (II) is useful in design and preparation of
CC ribozyme molecules for inhibiting expression of the tumour polypeptides
CC and proteins in tumour cells; and to isolate a full length gene from a
CC suitable library e.g., a tumour cDNA library using well known
CC techniques.
XX
SQ Sequence 496 BP; 107 A; 132 C; 157 G; 100 T; 0 other;
Query Match 25.8%; Score 450.8; DB 24; Length 496;
Best Local Similarity 96.8%; Pred. No. 3.4e-83;
Matches 482; Conservative 0; Mismatches 12; Indels 4; Gaps 2;

QY 203 CGCGGTCGCGGACCGCTGGCGCGCGCGGCGCGCGCGCTGGCGCTGAGTGCAGG 262
DB 2 CGCGCGGTCGCGGACCGCTGGCGCGCGCGCGCTGGCGCTGAGTGCAGTGCAGG 58
QY 263 CGGAGAGCTGCGGCTGCAGGAG-GAGAGCGTGGCGTGCACCGAGTTAACATCTACCTC 321

```

```

DB 59 GCAGGAGCTGGCGTGCAGGAGAGCGTGCAGTGCACGAGTTAACATCTACCTC 118
QY 322 AGCGACCGCATCTCACTGCACCGCGCTGGCGGAGCGCTGGAAACCGCTGTGCAAGAG 381
DB 119 AGCGACCGCATCTCACTGCACCGCGCTGGCGGAGCGCTGGAAACCGCTGTGCAAGAG 178
QY 382 AAGAAATATGATTATGATTAATTTGCCAGGACATCTGTTATCATAGCATTTTATATGAA 441
DB 179 AAGAAATATGATTATGATTAATTTGCCAGGACATCTGTTATCATAGCATTTTATATGAA 238
QY 442 GCCTGCTCAACTCTCTTCGGACAGTTTACAGTGTCTTTGAGACATCCCGGATATCCTG 501
DB 239 GCCTGCTCAACTCTCTTCGGACAGTTTACAGTGTCTTTGAGACATCCCGGATATCCTG 298
QY 502 CTAGAAGAAGTGTATCTTTGTAGATGACTACAGTGTAGAGAGCAGTGTGAGGAGCGCTTG 561
DB 299 CTAGAAGAAGTGTATCTTTGTAGATGACTACAGTGTAGAGAGCAGTGTGAGGAGCGCTTG 358
QY 562 GCCAATGAGCTTTTCGGGACTGCCCAAGGTGCGCTGATCCGCGCCCAACAGAGAGAGGGC 621
DB 359 GCCAATGAGCTTTTCGGGACTGCCCAAGGTGCGCTGATCCGCGCCCAACAGAGAGAGGGC 418
QY 622 CTGCTGCGAGCCCGCTGCTGGGGCGTCTGCGGAGGAGCGATGTTCTTGACCTTCTG 681
DB 419 CTGCTGCGAGCCCGCTTCTGGGGCGTCTGCGGCGAAGGGCGATGTTCTTGACCTTCTG 478
QY 682 GACTGTCACTGTGAGTGC 699
DB 479 GACTGTCACTGTGAGTGC 496

```

## RESULT 14

ABL81811/c  
ID ABL81811 standard; cDNA; 473 BP.

```

XX AC ABL81811;
XX
DT 17-MAY-2002 (first entry)
XX
DE Human ovarian cancer related cDNA clone SEQ ID NO:4789.
XX
KW Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200192581-A2.
XX
PD 06-DEC-2001.
XX
PF 29-MAY-2001; 2001WO-US17756.
XX
PR 26-MAY-2000; 2000US-207484P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Algate PA, Harlocker SL, Jones R;
XX
WPI; 2002-122075/16.
XX
Composition for therapy and diagnosis of ovarian cancer comprising
PT polypeptide of a ovarian tumor polypeptide, polynucleotide encoding
PT polypeptide, antibody specific to polypeptide or T cell expressing
PT polypeptide -
XX
Claim 1; SEQ ID 4789; 489pp; English.
XX
The present invention describes a composition (I) comprising: carriers
CC and immunostimulants; and a polypeptide (II) of a ovarian tumour
CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence
CC (S1) from the 10912 nucleotide sequences as given in ABL77023 to
CC ABL87934, (III) encoding (II) having a sequence (S2), a T cell
CC population of (II), or antigen presenting cells that express (II).

```



CC (I) has cytostatic activity. An oligonucleotide (IV) that hybridises to  
CC (SI) can be used for detecting ovarian cancer in a patient's biological  
CC sample preferably serum or ovarian tissue. The method comprises  
CC contacting a biological sample from a patient with (IV), detecting the  
CC amount of polynucleotide hybridising to (IV) and comparing the amount to  
CC a predetermined cutoff value and thereby detecting ovarian cancer in the  
CC patient, where the amount of polynucleotide hybridising to (IV) is  
CC detected preferably by polymerase chain reaction (PCR). (I) comprising  
CC (iii) and/or (ii) is useful for stimulating and/or expanding T cells  
CC specific for an ovarian tumour comprising contacting T cells  
CC with (iii) or (ii). (iii) is useful in design and preparation of  
CC ribozyme molecules for inhibiting expression of the tumour polypeptides  
CC and proteins in tumour cells; and to isolate a full length gene from a  
CC suitable library e.g., a tumour cDNA library using well known  
CC techniques.

XX SQ Sequence 473 BP; 108 A; 164 C; 103 G; 98 T; 0 other;

Query Match 25.4%; Score 443.4; DB 24; Length 473;  
Best Local Similarity 99.4%; Pred. No. 1.1e-81;  
Matches 466; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

QY 595 CTGATCCGCGCCCAACAGAGAGAGGGCTGTGCGAGCCCGCTGCTGGGGCGCTCTGCG 654  
DB 473 CTTATCCGCGCCCAACAGAGAGAGGGCTGTGCGAGCCCGCTGCTGGGGCGCTCTGCG 414  
QY 655 GCGAGGGCGGATGTTCTGACCTTCTGAGTGTCACTGTGAGTCCACAGAGGGTGGCTG 714  
DB 413 GCGAGGGCGGATGTTCTGACCTTCTGAGTGTCACTGTGAGTCCACAGAGGGT-GCTG 355  
QY 715 GAGCGCTGCTGCAGAGATCCATGAAGAGAGTCCGCGAGTGGTGTGCCCGGTGATTGAT 774  
DB 354 GAGCGCTGCTGCAGAGATCCATGAAGAGAGTCCGCGAGTGGTGTGCCCGGTGATTGAT 295  
QY 775 GTGATCGACTGGAAACACTTTCGAATACCTGGGAACTCCGGGAGCCCGAGATCGCGGT 834  
DB 294 GTGATCGACTGGAAACACTTTCGAATACCTGGGAACTCCGGGAGCCCGAGATCGCGGT 235  
QY 835 TTCGACTGGAGGCTGGTGTTCACG-TGGCACACAGTTCCTGAGAGGAGAGATACCGAT 893  
DB 234 TTCGACTGGAGGCTGGTGTTCACGCTGGCACACAGTTCCTGAGAGGAGAGATACCGAT 175  
QY 894 GCATCCCGCTGCATGTCATCAGTCTCCACAAATGGCTGGTGGCTGTTGCTGTGAG 953  
DB 174 GCATCCCGCTGCATGTCATCAGTCTCCACAAATGGCTGGTGGCTGTTGCTGTGAG 115  
QY 954 TAAGAAATATTTTGAATATCTGGGGTCTTATGATACAGGAATGGAAGTTTGGGAGGAGA 1013  
DB 114 TAAGAAATATTTTGAATATCTGGGGTCTTATGATACAGGAATGGAAGTTTGGGAGGAGA 55  
QY 1014 AAACCTCGAATTTTCTTTAGGATCTGGCAGTGTGGTGGGGTCTCTGGAA 1062  
DB 54 AAACCTCGAATTTTCTTTAGGATCTGGCAGTGTGGTGGGGTCTCTGGAA 6

## RESULT 15

ABL80705/c

XX ID ABL80705 standard; cDNA; 473 BP.

XX AC ABL80705;

XX DT 17-MAY-2002 (first entry)

XX DE Human ovarian cancer related cDNA clone SEQ ID NO:3683.

XX KW Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.

XX OS Homo sapiens.

XX PN W0200192581-A2.

XX PD 06-DEC-2001.

PF 29-MAY-2001; 2001WO-US17756.

XX PR 26-MAY-2000; 2000US-207484P.

XX PA (CORI-) CORIXA CORP.

XX PI Algate PA, Harlocker SL, Jones R;

XX DR WPI; 2002-122075/16.

XX PT Composition for therapy and diagnosis of ovarian cancer comprising

XX PT polypeptide of an ovarian tumor polypeptide, polynucleotide encoding

XX PT polypeptide, antibody specific to polypeptide or T cell expressing

XX PT polypeptide

XX PS Claim 1; SEQ ID 3683; 489pp; English.

XX PS The present invention describes a composition (I) comprising: carriers

XX CC and immunostimulants; and a polypeptide (II) of an ovarian tumour

XX CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence

XX CC (SI) from the 10912 nucleotide sequences as given in ABL77023 to

XX CC ABL87934. (III) encoding (II) having a sequence (S2), a T cell

XX CC population of (II), or antigen presenting cells that express (II).

XX CC (I) has cytostatic activity. An oligonucleotide (IV) that hybridises to

XX CC (SI) can be used for detecting ovarian cancer in a patient's biological

XX CC sample preferably serum or ovarian tissue. The method comprises

XX CC contacting a biological sample from a patient with (IV), detecting the

XX CC amount of polynucleotide hybridising to (IV) and comparing the amount to

XX CC a predetermined cutoff value and thereby detecting ovarian cancer in the

XX CC patient, where the amount of polynucleotide hybridising to (IV) is

XX CC detected preferably by polymerase chain reaction (PCR). (I) comprising

XX CC (iii) and/or (ii) is useful for stimulating and/or expanding T cells

XX CC specific for an ovarian tumour protein comprising contacting T cells

XX CC with (iii) or (ii). (iii) is useful in design and preparation of

XX CC ribozyme molecules for inhibiting expression of the tumour polypeptides

XX CC and proteins in tumour cells; and to isolate a full length gene from a

XX CC suitable library e.g., a tumour cDNA library using well known

XX CC techniques.

XX SQ Sequence 473 BP; 106 A; 162 C; 105 G; 100 T; 0 other;

Query Match 24.4%; Score 425.8; DB 24; Length 473;

Best Local Similarity 98.7%; Pred. No. 4.5e-78;

Matches 471; Conservative 0; Mismatches 2; Indels 4; Gaps 4;

QY 592 CGCTGATCCGCGCCCAACAGAGAGAGGGCTGTGCGAGCCCGCTGCTGGGGCGCTCT 651  
DB 473 CGCTGATCCGCGCCCAACAGAGAGAGGGCTGTGCGAGCCCGCTGCTGGGGCGCTCT 414  
QY 652 GCGCGAGGGCGGATGTTCTGACCTTCTGGAATGTCAGTGTGAGTCCACGAGAGGTGG 711  
DB 413 GCGCGGA-GGGCGATGTTCTGACCTTCTGGAATGTCAGTGTGAGTCCACGAGAGGT-G 356  
QY 712 CTGGAGCGCTGCTGCAGAGGATCCATGAAGAGAGTCCGCACTGTGTGTCGCCGGTATT 771  
DB 355 CTGGAGCGCTGCTGCAGAGGATCCATGAAGAGAGTCCGCACTGTGTGTCGCCGGTATT 296  
QY 772 GATGATCGACTGGAAACCTTCGAATACCTGGGAACTCCCGGGAGCCCGAGATCGGC 831  
DB 295 GATGATCGACTGGAAACCTTCGAATACCTGGGAACTCCCGGGAGAGATCGGC 237  
QY 832 GGTTCGACTGGAGGCTGGTGTTCAGTGTGCGACACAGATTCCTGAGAGGAGAGATACGG 891  
DB 236 GGTTCGACTGGAGGCTGGTGTTCAGTGTGCGACACAGATTCCTGAGAGGAGAGATACGG 177  
QY 892 ATGCAATCCCGCTCGATGTCAGGTCTTCCAAATATGCTGCTGGTGGTGTGCTGTG 951  
DB 176 ATGCAATCCCGCTCGATGTCAGGTCTTCCAAATATGCTGCTGGTGGTGTGCTGTG 117  
QY 952 AGTAAGAAATATTTTGAATATCTGGGGTCTTATGATACAGGAATGGAAGTTTGGGAGGA 1011  
DB 116 AGTAAGAAATATTTTGAATATCT-GGGTCTTATGATACAGGAATGGAAGTTTGGGAGGA 58



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 7, 2003, 12:45:48 ; Search time 6368.95 Seconds  
(without alignments)  
11215.070 Million cell updates/sec

Title: US-10-074-527-3

Perfect score: 1746

Sequence: 1 atgtgggggcacgcgcg.....tcaaagagcagtgattatga 1746

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 288711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_hvg.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_ats.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vt.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_om.\*

21: em\_or.\*

22: em\_ov.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pl.\*

26: em\_ro.\*

27: em\_ats.\*

28: em\_un.\*

29: em\_vi.\*

30: em\_htg\_hum.\*

31: em\_htg\_inv.\*

32: em\_htg\_other.\*

33: em\_htg\_mus.\*

34: em\_htg\_pln.\*

35: em\_htg\_rod.\*

36: em\_htg\_mam.\*

37: em\_htg\_vrt.\*

38: em\_sy.\*

39: em\_htgo\_hum.\*

40: em\_htgo\_mus.\*

41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1746	100.0	2745	6	AX477705
2	1744.4	99.9	1746	9	AB078146 Homo sapi
3	1744.4	99.9	1746	9	AB078146 Homo sapi
4	1319	75.5	2343	9	AK024865 Homo sapi
5	851.4	48.8	63632	2	Continuation (5 of
c	851.4	48.8	110000	2	Continuation (3 of
7	851.4	48.8	288118	2	Continuation (8 of
8	842.6	48.3	228368	2	Continuation (8 of
c	841	48.2	94616	2	Continuation (8 of
10	841	48.2	288118	2	Continuation (8 of
11	830	47.5	1928	9	BC013945 Homo sapi
12	812.6	46.5	110000	2	Continuation (4 of
c	812.6	46.5	166547	2	Continuation (4 of
14	805.4	46.1	166547	2	Continuation (4 of
c	803.8	46.0	110000	2	Continuation (2 of
16	795.4	45.6	110000	2	Continuation (2 of
c	795.4	45.6	110000	2	Continuation (2 of
18	795.4	45.6	299134	2	Continuation (3 of
c	792.2	45.4	110000	2	Continuation (2 of
21	526.2	30.1	328901	2	Continuation (2 of
22	522.4	29.9	226623	2	Continuation (2 of
c	520.8	29.8	1737	10	U73819 Mus muscu
24	520.8	29.8	234109	2	Continuation (2 of
c	510.6	29.2	110000	2	Continuation (2 of
26	499.4	28.6	5369	9	BC036390 Homo sapi
c	499.4	28.6	73481	9	AC025034 Homo sapi
28	496.2	28.4	1737	6	AR236647 Sequence
c	493.6	28.3	165175	2	Y08564 Homo sapien
31	419	24.0	554	6	AX197983 Sequence
32	418.4	24.0	544	6	AX208574 Sequence
c	400.8	23.0	652	6	AX197725 Sequence
34	400.2	22.9	632	6	AX208332 Sequence
35	400.2	22.9	632	6	AX209583 Sequence
36	374	21.4	149930	9	AL136084 Human DNA
37	363	20.8	483	6	AX371123 Sequence
c	300.8	17.2	2384	3	AK112431 Ciona int
39	285.6	16.4	1920	6	AR153422 Sequence
40	285.6	16.4	2575	6	AX376296 Sequence
c	285.6	16.4	2575	6	AX697278 Sequence
42	285.6	16.4	3357	9	AY035399 Homo sapi
43	285.6	16.4	3580	9	AL831925 Homo sapi
44	284	16.3	3332	6	AR155987 Sequence
45	284	16.3	3528	9	BC014789 Homo sapi

# ALIGNMENTS

RESULT 1  
AX477705  
LOCUS AX477705  
DEFINITION Sequence 17 from Patent WO0246426.  
ACCESSION AX477705  
VERSION AX477705.1 GI:22216860  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Sanjanwala,M.M., Yao,M.G., Au-Young,J., Baughn,M.R., Arvizu,C.,  
Ring,H.Z., Lee,E.A., Ding,L., Hafalia,A.J., Tang,Y.T., Yue,H.,  
Tribouley,C.M., Lu,D.A., Warren,B.A., Yang,J.,

2745 bp DNA linear PAT 12-AUG-2002

Walia,N.K., Nguyen,D.B., Gandhi,A.R. and Ison,C.H.

TITLE Drug metabolizing enzymes  
JOURNAL Patent: WO 0246426-A 17.13-JUN-2002;  
Incyte Genomics, Inc. (US)

FEATURES  
Location/Qualifiers  
1..2745

/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/note="Incyte ID No: 2860635CB1"

BASE COUNT 704 a 612 c 737 g 692 t

ORIGIN

Query Match	100.0%;	Score 1746;	DB 6;	Length 2745;
Best Local Similarity	100.0%;	Pred. No. 2e-266;		
Matches 1746;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	ATGTGGGGGCGACGGCGCGCGCGCTGCGCGGGAACTGCGCGCGCGGGAGGCG	60	
Db	1	ATGTGGGGGCGACGGCGCGCGCGCTGCGCGGGAACTGCGCGCGCGGGAGGCG	60	
Qy	61	CTGTTGGTGTCTTGGCGCTACTGGCGTTGGCGCGCGCTGGCGTGGCGTGGCG	120	
Db	61	CTGTTGGTGTCTTGGCGCTACTGGCGTTGGCGCGCGCTGGCGTGGCGTGGCG	120	
Qy	121	CGTGGGCGGGCG	180	
Db	121	CGTGGGCGGGCG	180	
Qy	181	CGCGAGCGCGTATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	240	
Db	181	CGCGAGCGCGTATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	240	
Qy	241	CGCGTGGCGTGCAGTGTGCGAGGCGAGAGCTGCGCGTGCAGAGAGAGCGT	300	
Db	241	CGCGTGGCGTGCAGTGTGCGAGGCGAGAGCTGCGCGTGCAGAGAGAGCGT	300	
Qy	301	CACGAGATTACATCTACTCAGCGAGCGCATCTCAGTGCACCGCGCGCGCG	360	
Db	301	CACGAGATTACATCTACTCAGCGAGCGCATCTCAGTGCACCGCGCGCGCG	360	
Qy	361	TGGAAACCGCTGTGCAAGAGAGAAATATGATTATGATATATGTTGCCAG	420	
Db	361	TGGAAACCGCTGTGCAAGAGAGAAATATGATTATGATATATGTTGCCAG	420	
Qy	421	ATCATAGCATTTTATATGAGAGCTGTGCTCTCTCTCTCTCTCTCTCTCT	480	
Db	421	ATCATAGCATTTTATATGAGAGCTGTGCTCTCTCTCTCTCTCTCTCTCT	480	
Qy	481	GAGACATCCCGGATATCTCTAGAGAGAGTATCTCTGTAGATGATGATGAT	540	
Db	481	GAGACATCCCGGATATCTCTAGAGAGAGTATCTCTGTAGATGATGATGAT	540	
Qy	541	GAGACCTGGAAGAGCGCTTGGCCAAATGAGCTTTTGGGACTTCCCAAGT	600	
Db	541	GAGACCTGGAAGAGCGCTTGGCCAAATGAGCTTTTGGGACTTCCCAAGT	600	
Qy	601	CGCGCCAAACAGAGAGAGCGCTGTGCGAGCGCGCGCTGTGCGCGCGAG	660	
Db	601	CGCGCCAAACAGAGAGAGCGCTGTGCGAGCGCGCGCTGTGCGCGCGAG	660	
Qy	661	GGCGATGTTCTGACCTTCTGAGCTGCTCAGTGTGAGTGCACGAGGCGT	720	
Db	661	GGCGATGTTCTGACCTTCTGAGCTGCTCAGTGTGAGTGCACGAGGCGT	720	
Qy	721	CTGCTGCAGAGGATCCATGAAGAGAGTCTGGCGAGTGTGTCGCCGCTG	780	
Db	721	CTGCTGCAGAGGATCCATGAAGAGAGTCTGGCGAGTGTGTCGCCGCTG	780	
Qy	781	GACTGGAAACACCTTGGAAATACCTGGGGAACTCCCGGGAGCGCCAGAT	840	
Db	781	GACTGGAAACACCTTGGAAATACCTGGGGAACTCCCGGGAGCGCCAGAT	840	

Qy	841	TGGAGGCTGTGTTCACGTGGCACACAGTTCTTGAGAGGAGGAGATACGGATGCAATCC	900	
Db	841	TGGAGGCTGTGTTCACGTGGCACACAGTTCTTGAGAGGAGGAGATACGGATGCAATCC	900	
Qy	901	CCCGTGCATGTCATCAGGTCTCCAAACAATGGCTGTGGGCTGTTGCTGTGAGTAAGAA	960	
Db	901	CCCGTGCATGTCATCAGGTCTCCAAACAATGGCTGTGGGCTGTTGCTGTGAGTAAGAA	960	
Qy	961	TATTTTGAATATCTGGGCTTATGATACAGGAATGGAAGTTTGGGGAGGAGAAACCTC	1020	
Db	961	TATTTTGAATATCTGGGCTTATGATACAGGAATGGAAGTTTGGGGAGGAGAAACCTC	1020	
Qy	1021	GAATTTTCTTTAGGATCTGCGAGTGTGGGCTTCTGGAACAACACCCATGTTCCCAT	1080	
Db	1021	GAATTTTCTTTAGGATCTGCGAGTGTGGGCTTCTGGAACAACACCCATGTTCCCAT	1080	
Qy	1081	GTTGGGCAATGTTTCCCAAGCAAGCTCCCTACTCCCGCAACAAGGCTCTGGCCAAAGT	1140	
Db	1081	GTTGGGCAATGTTTCCCAAGCAAGCTCCCTACTCCCGCAACAAGGCTCTGGCCAAAGT	1140	
Qy	1141	GTTGCTGCAGCTGAAGTATGATGGATGGAATTTAAAGAGCTCTACTACCATCGCAACCCC	1200	
Db	1141	GTTGCTGCAGCTGAAGTATGATGGATGGAATTTAAAGAGCTCTACTACCATCGCAACCCC	1200	
Qy	1201	CGTGGCCGCTTGGAAACCTTTTGGGATGTGACAGAGAGGAGAGCTCCGGGACAAAGCTC	1260	
Db	1201	CGTGGCCGCTTGGAAACCTTTTGGGATGTGACAGAGAGGAGAGCTCCGGGACAAAGCTC	1260	
Qy	1261	CAGTGTAAAGACTTCAAGTGGTCTTGGAGACTGTGTATCCAGAACTGTCATGTCCTGAG	1320	
Db	1261	CAGTGTAAAGACTTCAAGTGGTCTTGGAGACTGTGTATCCAGAACTGTCATGTCCTGAG	1320	
Qy	1321	GACAGGCTGGCTCTTCCGGATGCTCCAGAAACAAAGGACTTAACAGACTACTGCTTTGAC	1380	
Db	1321	GACAGGCTGGCTCTTCCGGATGCTCCAGAAACAAAGGACTTAACAGACTACTGCTTTGAC	1380	
Qy	1381	TATAACCTCCCGATGAAAACCCAGATTGTGGGACACACAGGTCATTCTGTACCTCTGTCAT	1440	
Db	1381	TATAACCTCCCGATGAAAACCCAGATTGTGGGACACACAGGTCATTCTGTACCTCTGTCAT	1440	
Qy	1441	GGGATGGGCGCAGAACTAGTTTTTCAGTACAGTCCAGAAAGAAATACGCTATAACACC	1500	
Db	1441	GGGATGGGCGCAGAACTAGTTTTTCAGTACAGTCCAGAAAGAAATACGCTATAACACC	1500	
Qy	1501	CACGAGCTGAGGCTGCAATGCTGTGGAAGAGGATGATACCTTATCATCATCTCTC	1560	
Db	1501	CACGAGCTGAGGCTGCAATGCTGTGGAAGAGGATGATACCTTATCATCATCTCTC	1560	
Qy	1561	TGCGAAGAACTGCCCCAGAGAAATCAGAAAGTTTCACTTTGAGGAGGATGGATCTTTATTT	1620	
Db	1561	TGCGAAGAACTGCCCCAGAGAAATCAGAAAGTTTCACTTTGAGGAGGATGGATCTTTATTT	1620	
Qy	1621	CACGAAACAGTCCAAAGAAATGTGTCAGGCTGCGAGGAGGAGTGCAGTGCAGTTTCGTT	1680	
Db	1621	CACGAAACAGTCCAAAGAAATGTGTCAGGCTGCGAGGAGGAGTGCAGTGCAGTTTCGTT	1680	
Qy	1681	CCACTCTTACGAGACTGCACCACTCGGATCATCAGAAATGGTTCTTCAAGAGCGCATG	1740	
Db	1681	CCACTCTTACGAGACTGCACCACTCGGATCATCAGAAATGGTTCTTCAAGAGCGCATG	1740	
Qy	1741	TTATGA 1746		
Db	1741	TTATGA 1746		

RESULT 2

AB078146

LOCUS

DEFINITION

AB078146

1746 bp

mRNA

linear

PRI 04-SEP-2002

Homo sapiens

GALNT12 mRNA for

UDP-N-acetyl-alpha-D-galactosamine:polypeptide

N-acetylglucosaminyltransferase 12, complete cds.

AB078146

AB078146.1

GI:22122073

[illegible]

Db 1321 CAGGTAAAGACTTCAAGTGGTCTTGAGACTGTGTATCCAGAACTGCATGTGCTGAG 1320

Qy 1321 GACAGGCTGGCTTCTTGGGATGCTCCAGAACAAAGACTTAACAGACTACTGCTTTGAC 1380

Db 1321 GACAGGCTGGCTTCTTGGGATGCTCCAGAACAAAGACTTAACAGACTACTGCTTTGAC 1380

Qy 1381 TATAACCTCCCGATGAAACACAGATTGTGGACACACAGGTCAATTCTGTACCTCTGTGCAT 1440

Db 1381 TATAACCTCCCGATGAAACACAGATTGTGGACACACAGGTCAATTCTGTACCTCTGTGCAT 1440

Qy 1441 GGGATGGGCCAGAAATCAGTTTTTTCAGGTACACGTCCAGAAAGAAATACGCTATAACACC 1500

Db 1441 GGGATGGGCCAGAAATCAGTTTTTTCAGGTACACGTCCAGAAAGAAATACGCTATAACACC 1500

Qy 1501 CACAGCTGAGGCTGCAATGCTGTGGAGCAGCAATGGAATACCTTATCATGCTC 1560

Db 1501 CACAGCTGAGGCTGCAATGCTGTGGAGCAGCAATGGAATACCTTATCATGCTC 1560

Qy 1561 TGGAGAGAACTGCCAGAGAAATCAGAAATTCATCTTGCAGGAGGATGGATCTTTATTT 1620

Db 1561 TGGAGAGAACTGCCAGAGAAATCAGAAATTCATCTTGCAGGAGGATGGATCTTTATTT 1620

Qy 1621 CACGAACAGTCCAGAAATGTGTCCAGCTCCGAGGAGGATCGAGTGACAGTTTCCTT 1680

Db 1621 CACGAACAGTCCAGAAATGTGTCCAGCTCCGAGGAGGATCGAGTGACAGTTTCCTT 1680

Qy 1681 CCATCTTACAGACTGCACCACTCGGATCATCAGAAATGGTTCTTCAAGAGCGCATG 1740

Db 1681 CCATCTTACAGACTGCACCACTCGGATCATCAGAAATGGTTCTTCAAGAGCGCATG 1740

Qy 1741 TTATGA 1746

Db 1741 TTATGA 1746

RESULT 3

HSAL132365

LOCUS HSAL132365 1746 bp mRNA linear PRI 27-NOV-2002

DEFINITION Homo sapiens mRNA for UDP-GalNAC-transferase 12 (GALNT12 gene).

ACCESSION AJ132365

VERSION AJ132365.2 GI:25815115

KEYWORDS GALNT12 gene; UDP-GalNAC-transferase 12.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Bennett,E.P.

Direct Submission

Submitted (20-JAN-1999) Bennett E.P., Glycobiology Group - 24.5.33, School of dentistry, Noerre Alle' 20, Copenhagen 2200N, DENMARK

REMARK revised by author [25-NOV-2002]

FEATURES

Location/Qualifiers

1..1746

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

1..1746

/gene="GALNT12"

1..1746

/gene="GALNT12"

/codon\_start=1

/product="UDP-GalNAC-transferase 12"

/protein\_id="CAC80100.2"

/db\_xref="GI:25815116"

/translation="MWGTRARRRCPRELRRGREALLVLLALLALAGLSVLRAQAGAGAAEPGPRPRPGRPEVMPRPVPAALARGAEAVRLQLOQEELRLQESVRLHQ INIYLSRISLHRLRPERNPLCKEKYDYNLPTSVIIAIFYNEASTLRLTYSVL ETPDILLEEVLVDYSDREHLKERLANELSGLPVKVLRIRANKREGVLRLLGASA ARGVLLTFLDCHECEGWLEPLLRITHEESA VVCPVIDVDWNTFEYLSNGEPQI GGFWRLVFTWHTYPERIRIRMQSPDVIRPWRAGLFAVSKRYFEYLSGDEMEF WGENLEFSFRITWCGVLETHPCSHVGHVFPKQAPYSRNKALANSVRAAEVWMDPEF ELYLHRNPRARLEPPGDVTERKQLRDLQCKDFKWFLETVVPELVHPEDRPGFTGMQLQ

NRKLTLYCFDYNPDPENQIVGHVILYLCHGMGQNOFFBYTQKEIRYNTHOPEGCIA  
VSAGMDTLIMHLCEETAPENQKFILODGLSFHEQSKVCQAARKSSDSFVPLLRDC  
TNSDQKWFKKERML"

BASE COUNT 382 a 449 c 549 g 366 t

ORIGIN

Query Match 99.9%; Score 1744.4; DB 9; Length 1746;  
Best Local Similarity 99.9%; Pred. No. 3.8e-266;  
Matches 1745; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGTGGGGCGCACGCGCGCGCGCTGCTGCCCGCGGAACTGCGGCGCGCGCGGAGGCG 60

Db 1 ATGTGGGGCGCACGCGCGCGCGCTGCTGCCCGCGGAACTGCGGCGCGCGCGGAGGCG 60

Qy 61 CTGTTGCTGCTCTGCGGCTTACTGCGGCTTGGCGGCGCTGCGGCTCGCTGCTGCGGCGCAG 120

Db 61 CTGTTGCTGCTCTGCGGCTTACTGCGGCTTGGCGGCGCTGCGGCTCGGCTCGGCTGCGGCGCAG 120

Qy 121 CGTGGGGCGCGGGCGCGGGGCTGCCGAGCCGGGACCCCGCGCACCCCGCGCCCCCGGCGG 180

Db 121 CGTGGGGCGCGGGCGCGGGGCTGCCGAGCCGGGACCCCGCGCACCCCGCGCCCCCGGCGG 180

Qy 181 CGCGAGCGGTCTATGCGCGCGCGCGCGGTGCGCGGAAACGCGCTGGCGCGCGCGGCGGAG 240

Db 181 CGCGAGCGGTCTATGCGCGCGCGCGCGGTGCGCGGAAACGCGCTGGCGCGCGCGGCGGAG 240

Qy 241 GCGGTGCGGCTGCGAGCTGCGAGGCGGAGGAGCTGCGGCTGCGAGGAGAGAGCGTGCAGCTG 300

Db 241 GCGGTGCGGCTGCGAGCTGCGAGGCGGAGGAGCTGCGGCTGCGAGGAGAGAGCGTGCAGCTG 300

Qy 301 CACGAGATTAACTTACCTCAGCGACCGCATCTCACTGCACCGCGCGCTGCCCGAGCGC 360

Db 301 CACGAGATTAACTTACCTCAGCGACCGCATCTCACTGCACCGCGCGCTGCCCGAGCGC 360

Qy 361 TGGAAACCGCTGTGCAAAAGAGAAATATGATTATGATTAATTTGCCAGAGCATCTGTT 420

Db 361 TGGAAACCGCTGTGCAAAAGAGAAATATGATTATGATTAATTTGCCAGAGCATCTGTT 420

Qy 421 ATCATAGCATTTTATATGAAAGCTTGTCAACTCTCTTCGGACAGTTTACAGTGTCTT 480

Db 421 ATCATAGCATTTTATATGAAAGCTTGTCAACTCTCTTCGGACAGTTTACAGTGTCTT 480

Qy 481 GAGACATCCCGGATATCTCTGTAGAAAGTGTATCTTGTAGATGACTACAGTGATAGA 540

Db 481 GAGACATCCCGGATATCTCTGTAGAAAGTGTATCTTGTAGATGACTACAGTGATAGA 540

Qy 541 GAGCACCCTGAAGGAGCGCTTGGCCAATAGCTTTGGGACTGCCCAAGAGTGCCTGATC 600

Db 541 GAGCACCCTGAAGGAGCGCTTGGCCAATAGCTTTGGGACTGCCCAAGAGTGCCTGATC 600

Qy 601 CGCGCCCAAGAGAGAGGCGCTGTCGAGCGCGGCTGTCGGGGCGCTCTGGCGCGAGG 660

Db 601 CGCGCCCAAGAGAGAGGCGCTGTCGAGCGCGGCTGTCGGGGCGCTCTGGCGCGAGG 660

Qy 661 GCGCATCTTCTGACCTTCTCTGAGTGTCACTGTGAGTGCCACCAAGAGGTGGTGAGCCG 720

Db 661 GCGCATCTTCTGACCTTCTCTGAGTGTCACTGTGAGTGCCACCAAGAGGTGGTGAGCCG 720

Qy 721 CTGCTGAGAGGATCCATGAAGAGGAGTCCGGAGTGTGTGTCGCCGCTGATGTGATC 780

Db 721 CTGCTGAGAGGATCCATGAAGAGGAGTCCGGAGTGTGTGTCGCCGCTGATGTGATC 780

Qy 781 GACTGGAAACCTTCGAAATACCTGGGAACTCCGGGAGCGCCAGATCGCGGTTTCGAC 840

Db 781 GACTGGAAACCTTCGAAATACCTGGGAACTCCGGGAGCGCCAGATCGCGGTTTCGAC 840

Qy 841 TGGAGGCTGTGTTCAGCTGGCAACAGTTCCTGAGAGGAGAGGATACGGATCAATCC 900

Db 841 TGGAGGCTGTGTTCAGCTGGCAACAGTTCCTGAGAGGAGAGGATACGGATCAATCC 900

Qy 901 CCGCTCGATGCTACAGGTCTCCAAATGGTGGTGGCTGTTTGTGTGAGTAAGAAA 960

Db 901 CCGCTCGATGCTACAGGTCTCCAAATGGTGGTGGCTGTTTGTGTGAGTAAGAAA 960



Qy	848	TG	TGTTT	CA	CGTGG	CA	CAC	AGTTT	CCT	GAGAGG	GAGAGG	AT	AC	GGAT	AC	CGAT	TG	CA	AT	CCCCCG	TG	907	
Db	421	TG	TGTTT	CA	CGTGG	CA	CAC	AGTTT	CCT	GAGAGG	GAGAGG	AT	AC	GGAT	AC	CGAT	TG	CA	AT	CCCCCG	TG	480	
Qy	908	AT	GT	CAT	CAG	CT	CC	AA	CA	AT	GG	CT	GG	CT	GGT	TG	TG	TG	TG	TG	TG	967	
Db	481	AT	GT	CAT	CAG	CT	CC	AA	CA	AT	GG	CT	GG	CT	GGT	TG	TG	TG	TG	TG	TG	540	
Qy	968	AA	TAT	CT	GG	GGT	CT	TT	AT	TAC	AG	GA	AT	GG	AA	GT	TT	GG	GG	GAG	AG	1027	
Db	541	AA	TAT	CT	GG	GGT	CT	TT	AT	TAC	AG	GA	AT	GG	AA	GT	TT	GG	GG	GAG	AG	600	
Qy	1028	CC	TTT	AG	AG	TCT	GG	CA	GT	GG	GGT	TCT	GG	AA	CA	CA	CC	AT	GT	CC	AT	1087	
Db	601	CC	TTT	AG	AG	TCT	GG	CA	GT	GG	GGT	TCT	GG	AA	CA	CA	CC	AT	GT	CC	AT	660	
Qy	1088	AT	GT	TTT	CC	CA	AG	CA	GC	T	CC	CT	CC	GG	CA	CA	AG	GC	T	GG	CA	1147	
Db	661	AT	GT	TTT	CC	CA	AG	CA	GC	T	CC	CT	CC	GG	CA	CA	AG	GC	T	GG	CA	720	
Qy	1148	C	AG	CT	GA	GT	AT	CG	AT	CG	AT	GA	TT	AA	A	G	AG	CT	T	AC	T	1207	
Db	721	C	AG	CT	GA	GT	AT	CG	AT	CG	AT	GA	TT	AA	A	G	AG	CT	T	AC	T	780	
Qy	1208	G	CT	T	G	GA	A	C	T	T	T	G	G	G	A	T	G	A	G	A	G	A	1267
Db	781	G	CT	T	G	GA	A	C	T	T	T	G	G	G	A	T	G	A	G	A	G	A	840
Qy	1268	A	A	G	A	C	T	T	C	A	A	G	T	T	C	T	T	G	A	G	A	G	1327
Db	841	A	A	G	A	C	T	T	C	A	A	G	T	T	C	T	T	G	A	G	A	G	900
						</																	

1  
 2  
 3  
 4  
 5  
 6

## RESULT 5

AC132018

**WPCOMMENT**

CONTINUED

sequence split into 3 fragments locus

	. Fragment Name	Begin	End
--	-----------------	-------	-----

AC132018 0	1	110000
------------	---	--------

AC132018-1	100001	210000
------------	--------	--------

AC132018-1	100001	210000
AC132018-2	200001	310000

AC132018\_2 200001 310000 200001 410000

Query Match	48.8%;	Score 851.4;	DB 2;	Length 63632;
Best Local Similarity	79.7%;	Pred. No. 3.7e-125;		
Matches 1081;	Conservative	0;	Mismatches 265;	Indels 10; Gaps 6;
QY	388	TATGATTATGATAAATTTGCCAGGACATCTGTTATCATAGCATTTTATATGAAGCCTGG	447	
DB	44126	TATGATTATGATAACCTGCCAGACATCCGTTGTCAAGCATTTTATATGAAGCCTGG	44185	
QY	448	TCAACTCTCCTTTGGACAGTTTACAGTGTCTCTGAGACATCCCGGATATCCTCTAGAA	507	
DB	44186	TCCACACTCCTTCAGACAGTTTACAGTGTCTCGAGACTTCCCTCTGATATCCTCTGGAG	44245	
QY	508	GAAGTGATCTCTGTGTAGACTACAGTGATAGAGACACCTGAAGGAGCGCTTGGCCAAT	567	
DB	44246	GAGGTCAATTCGTGTAGATGACTACAGCGACAGAGATCACCTGAAGGAGCCCTGGTCAA	44304	
QY	568	GAGCTTTGGGAGTCGCCAAGGTGGCGCTGATCGCGCCAAAGAGAGAGGGGCTGGTG	627	
DB	44305	-GAGCTTTCACAGCTGCCCTAGGTGACCTGATCCTCTGTAGCA--AGAGAGGACTAGTG	44361	
QY	628	CGAGCCCGGCTGCTCGGGGGCGTCTCGCGCGAGGGGCGATGTTCTGACCTTCTCGACTGT	687	
DB	44362	CAAG--CGGCTACTGGGAGCCTTTGCGCGCAGGGGCGAAGTGCTGAGGTTTCTGGACTGT	44420	
QY	688	CACTGTGAGTGCCAGAAAGGTGGCTGAGCGCGTCTGCGAGAGATCCATGAAGAGGAG	747	
DB	44421	CACTGTGACTGTCAAGAGGGGTGGCTGGAGTCCCTGCTCGAGAGGATCCACGAGAAAGGAG	44480	
QY	748	TGGCAGTGGTGTGCCGGTGATTCATCTGATCGACTGGAACACCTTCGAATACCTGGGG	807	
DB	44481	TGCGGCGTGGTGTGCCGGGTTATCGATGTCAATGTACTGGAAACACCTCTGAGTACTCTGGGC	44540	
QY	808	AACTCCGGGAGCCCCAGATCGCGCGTTTCGACTGGAGGCTGGTGTTCACGTGGGCACACA	867	
DB	44541	AACTCCGGAGACCCCGAGATTGGTGGCTTTGACTGGCGCCTAGTATTACCTGGCAGGTG	44600	
QY	868	GTTCTGTAGAGGAGGAGGATACGGATGCAATCCCCCGTGCAGTGTCACTCAGGTCTCCAAACA	927	
DB	44601	GTCCCTTGGCGGAGCGGAAGTTGATCGGGACTCCATAATGACATTTATCAGGTCTCCAACT	44660	
QY	928	ATGGCTGGTGGGCTGTTTGCTGTGTAGTAGAATAATTTTGAATACTCGGGGCTTATGAT	987	
DB	44661	ATGGCTTGGGGGACTGTGTTGTGTAGTAGAGATATTTTGAATACCTGGGGTCTTATGAT	44720	
QY	988	ACAGGAATCGAAAGTTTGGGGAGGAGAAACCTCGAAATTTTCTTTAGGATCTGGCAGTGT	1047	
DB	44721	ACCGGAATGGAAGTCTGGGGAGGAGAAACCTTGAATTTCTCCTTAGGATCTGGCAGTGT	44780	
QY	1048	GGTGGGGTTCTGGAACAACACCCATGTTTCCCAATGTGGCCATGTTTTTCCCAAGCAAGCT	1107	
DB	44781	GGT--GACTCTGGAACAACACCTTTACTCCCAAGTGGGCGACGTTCTCCCTAAGCAAGCT	44838	
QY	1108	CCCTACTCCCGMAAAGGCTCTGGCCAAACAGTGTTCGTGCAGCTGAAGTATGGATGAT	1167	
DB	44839	CCCTATTACGCGAGCAAGGCTCTAGCCAAACAGTGTCCGAGCTGCAGAAAGTGTGGACGGAT	44898	
QY	1168	GAAATTTAAAGAGCTCTACTACCATTCCGAACCCCGTGGCGCTTGGAACTTTTGGGGAT	1227	
DB	44899	GAACTTTAAAGAACTCTACACACCCCGMAAACCC--AGCCCGCTGGAACTCTGGGAG	44957	
QY	1228	GTGACAGAGAGGAAGCAGCTCCGGGACAAGCTCCAGTGTAAAGACTTCAAGTGTGTTCTTG	1287	
DB	44958	CTGACAGAGAAAGAAGCTTCAGCCCAAGCTCCAGTGTGAAGACTTAAAGCGGTTCCTG	45017	
QY	1288	GAGACTGTGTATCCAGAACTGCATGTGCTGTAGGACAGGCCCTGGCTTCTTCGGGATGCTC	1347	
DB	45018	GCTACACTGTATCCAGAACTGCATGTGCCAGAGGACAGGCCCTGGTGTCTTCGGGAAGCTT	45077	
QY	1348	CAGAACAAAGGACTAACAGACTACTGCTTTTGTACTATAACCTTCCCGATGAAAAACAGATT	1407	
DB	45078	GAGAAACAGAGGACTACGGGGGTACTGCTGTGGACTATATCTCTCCAGTAGAACAATGTG	45137	
QY	1408	GTGGGACACAGGTCAATCTGTACTCTGTGATGGGATGGGCGCAATCAAGTATTTTCGAG	1467	



Db 45138 GAAGCCACAGGTCCTTTTGTACTCTGCCACGGATGGTCAAGAACCGATTTTTCGAG 45197  
Qy 1468 TACACGTCCAGAGAAAGTAACGCTATAACACACCCAGCCTGAGGGCTGCATTGCTGTG 1527  
Db 45198 TGTACATCCCGGCAAGAAATACGGTATAACACACCCGCGCAGCCCGAGGCTCTGATAGCAGTG 45257  
Qy 1528 GAAGCAGGAATGATACCTTTATCATGATCTCTGCGAAGAACTGCCGCCAGAGATCAG 1587  
Db 45258 GAGGAAGGGAAGGAT--CCTTATCCCGGATCTCTGGAAGATATCCGTTCCAGAGAAATCAA 45315  
Qy 1588 AAGTTTCATCTGACGAGGAGGATGATCTTTATTTTCAACGACAGTCCAAAGAAATGTGTCAG 1647  
Db 45316 GAGTTTCATCTACAGGAGGACGCGACGTTAGTCTTACAGCAGACGGGAAATGTGTCAG 45375  
Qy 1648 GCTGCGAGGAGGAGTCCGAGTGACAGTTTCGTTCCACTCTTACGAGACTGACCAACTCG 1707  
Db 45376 GCCACAGAGGAGGAGTTCAACTCTGCTCTGCTCCGAACTTTCGAGACTGTAGCAACTCA 45435  
Qy 1708 GATCATCAGAAATGGTTCTTCAAGAGGCGATGTTAT 1744  
Db 45436 GACAGCCAGAGTGTGCTCTTCAAGGAGCACATGTCTAT 45472

RESULT 6  
AC098559 2/c  
WPCOMMENT

Sequence split into 8 fragments LOCUS AC098559 Accession AC098559

Fragment Name	Begin	End
AC098559_0	1	110000
AC098559_1	100001	210000
AC098559_2	200001	310000
AC098559_3	300001	410000
AC098559_4	400001	510000
AC098559_5	500001	610000
AC098559_6	600001	710000
AC098559_7	700001	794616

Continuation (3 of 8) of AC098559 from base 200001 (AC098559 Rattus norvegicus clone CH2

Query Match 48.8%; Score 851.4; DB 2; Length 110000;  
Best Local Similarity 79.7%; Pred. No. 3.5e-125;  
Matches 1081; Conservative 0; Mismatches 266; Indels 10; Gaps 6;

Qy 388 TATGATTATGATTAATTTGCCAGACATCTGTTATCATAGATTTTATATGAAGCCTGG 447  
Db 95301 TATGATTATGATTAATTTGCCAGACATCTGTTATCATAGATTTTATATGAAGCCTGG 95242  
Qy 448 TCAACTCTCTCGGACAGTTTACAGTGTCTTGAGACATCCCGGATATCCTGCTAGAA 507  
Db 95241 TCCACACTCTCTCAGACAGTTTACAGTGTCTTGAGACTTCCCTGTATATCCTGCTGGAG 95182  
Qy 508 GAAGTGATCTCTGTAGATGACTACAGTGATAGAGACACCTGAAGGAGCGCTTGGCCAAT 567  
Db 95181 GAGTTCATTCTGTAGATGACTACAGCGACAGAGATCACCTGAAGGAGCGCTTGGTCAA - 95123  
Qy 568 GAGCTTTCCGAGCTGCCCAAGGTGCGCTGATCCGCCCAACAGAGAGAGGCGCTGGTG 627  
Db 95122 -GAGCTTTCACAGTGTCCCTAGGTGCACCTGATCCTGCTAGCA--AGAGAGGACCTAGTG 95066  
Qy 628 CGAGCCCGGCTGTGGGGGCGTCTGCGGCGAGGGCGATGTTCTGACCTTCTGGAAGTGT 687  
Db 95065 CAAG-CCGGTACTGCGGAGCGCTTTCGCGCCAGGGGCGAAGTGTCTGAGGTTTCTGGAAGTGT 95007  
Qy 688 CACTGTGAGTGCCACGAGGAGTGTGCTGAGCGCGTGTGTCAGAGGATCCATGAAGAGGAG 747  
Db 95006 CACTGTGAGTGTCTATGAGGAGTGTGCTGAGTGTCTGTCGAGAGATCCAGAGAGGAG 94947  
Qy 748 TCGGCAGTGTGTGCCCGGTGATTGATGATGATGATGATGATGATGATGATGATGATGATG 807  
Db 94946 TCGGCAGTGTGTGCCCGGTGATTGATGATGATGATGATGATGATGATGATGATGATGATG 807  
Qy 808 AACTCCGGGAGGCCCAAGATCGGGCGTTTGCATGAGGCTGTGTTACGTGTCACATCA 867

Db 94886 AACTCCGGAGAGCCCGAGATTGTTGCTTTGACTGGCGCCTAGTATTACGTGCGCAGGTG 94827  
Qy 868 GTTCTCTGAGAGGAGAGGATACCGATGCAATCCCGCGTGCATGTGTCATCAGGTCTCCAAACA 927  
Db 94826 GTCCCTGGCGGAGCGGAAGTTGATGCGGACTCCTATTGACATTTATCAGGTCTCCAACT 94767  
Qy 928 ATGCTGTGGTGGCTGTTTGTGCTGAGTAAGAAATATTTGAATATCTGGGGTCTTATGAT 987  
Db 94766 ATGCTGTGGGAGCTGTTTGTGTTGAGTAAGAGATATTTTGAATACCTGGGGTCTTATGAT 94707  
Qy 988 ACAGGAATGAAGTTTGGGAGGAGAAACCTCGAATTTTCTTTAGGATCTGCGAGTGT 1047  
Db 94706 ACCGGAATGGAATCTGGGAGGAGAAACCTTGAATTTCTCTTTAGGATCTGCGAGTGT 94647  
Qy 1048 GGTGGGTTCTGGAACACACACCCATGTTCCCATGTTTGGCCATGTTTTCCTCCCAAGCAAGCT 1107  
Db 94646 GGT--GACTCTGGAACACACACCCCTTACTCCACAGTGGGCCACGCTTTCCTTAAGCAAGCT 94589  
Qy 1108 CCTACTCCCGCAACAGGCTCTGGCCACACAGTGTTCGTGAGCTGAAGTATGATGAT 1167  
Db 94588 CCTATTTCACGACAGAGGCTCTAGCCACACAGTGTCCGAGCTGCAGAGTGTGACGAGT 94529  
Qy 1168 GAATTTAAAGAGCTCTACTACCATCGCAACCCCGTCCCGCTTGGAACTTTTGGGGAT 1227  
Db 94528 GAACTTAAAGAACTCTACCAACCCCGGAAACCCCG--AGCCCGCTTGGAACTTCTGGGGAG 94470  
Qy 1228 GTGACAGAGAGGAGAGCTCTCCGGGACACAGCTCCAGTGTAAAGACTTCAAGTGTCTTCTTG 1287  
Db 94469 CTGACAGAGAGAAAGAGCTTTCAGCCCAAGCTCCAGTGTGAAGACTTAAAGCGGTCTCTG 94410  
Qy 1288 GAGACTGTGTATCCAGAACTGCAATGTCCTGAGAGACAGGCTTGGCTTCTTCGGAGTGTCTC 1347  
Db 94409 GCTACACTGTATCCAGAACTGTCATGTGCCAGAGACAGGCTGCTTCTTCGGGAAGCTT 94350  
Qy 1348 CAGAACAAAGGACTTAACAGACTACTCTTGTGACTATTAACCTCCCGATGAAACCAAGATT 1407  
Db 94349 GAGAACAGAGGACTTACGGGGTACTGCTGGAATTAATCTCCAGTGAAACCAATGTG 94290  
Qy 1408 GTGGGACACAGAGTCTATCTGTACCTCTGTATGGGATGGCGGACAGATCAAGTCTTTCGAG 1467  
Db 94289 GAAGGCCACAGGCTCTCTTTGTACTCTGCGACGGATGGGTGAGAACCAAGTCTTTCGAG 94230  
Qy 1468 TACAGCTCCAGAAAGAAATACGCTATAACCCCAACAGCTGAGGCTGAGGCTGATGTGTG 1527  
Db 94229 TGTACATCCCGCAAGAAATACGGTATAACCCCGCAGCCAGGCTGCGATAGCAGTG 94170  
Qy 1528 GAAGCAGGATGATACCTTATCATGCTCTCTGCGAAGAACTGCCCGAGAAATCAG 1587  
Db 94169 GAGGAAGGGAAGGAT--CCTTATCCCGGATCTCTGGAAGATACCCGTTCCAGAGAAATCAA 94112  
Qy 1588 AAGTTCATCTTGCAGGAGGATGATCTTTATTTTCAAGACAGTCCAGAAATGTGTCAG 1647  
Db 94111 GAGTTCATCTTACAGGAGGACGACGCTTAGTCTACAGCAGAGCGGGAATGTGTGAG 94052  
Qy 1648 GCTGCGAGGAAGAGTGTGAGTGTGAGTGTGCTTCCACTCTTACGAGACTGCAACACTCG 1707  
Db 94051 GCCACAGAGGAGGAGTTCACCTCTGCTTGTCTCCGAACTTGGCAGACTGTAGCAACTCA 93992  
Qy 1708 GATCATCAGAAATGTTTCTTCAAGAGCGCATCTTAT 1744  
Db 93991 GACAGCCAGAGGTGGTCTTCAAGGAGCACATGTCTAT 93955

RESULT 7

AC129139

LOCUS

DEFINITION

AC129139

VERSION

KEYWORDS

SOURCE

ORGANISM

AC129139 288118 bp DNA linear HTG 11-OCT-2002  
Rattus norvegicus clone CH230-293111, \*\*\* SEQUENCING IN PROGRESS  
\*\*\*, 14 unordered pieces.  
AC129139  
AC129139 2 GI:23829088  
HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
Rattus norvegicus (Norway rat)  
Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 288118)

Muzny,D,Marie., Metzker,M, Lee., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, C., Allen, C., Alsebrook, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Blawie, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Haviak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacobs, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpachy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhewa, L., Loulseghe, H., Lozada, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokeme, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Snajds, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, D., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstein, G. and Gibbs, R.A.

# TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

Center: Genome Center  
Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

----- Project Information

Center project name: GRGM

Center clone name: CH230-293111

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 214989 bases at least Q40

Consensus quality: 22817 bases at least Q30

Consensus quality: 27320 bases at least Q20

Estimated insert size: 252471; sum-of-contigs estimation

Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html))

\* NOTE: This sequence may represent more than one clone.

\* NOTE: This is a 'working draft' sequence. It currently

consists of 14 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

-----

1 5778: contig of 5778 bp in length  
5878: gap of unknown length  
5879: contig of 10495 bp in length  
16374: gap of unknown length  
16473: gap of unknown length  
16474: contig of 73947 bp in length  
90421: gap of unknown length  
90520: gap of unknown length  
90521: contig of 146262 bp in length  
236783: gap of unknown length  
236882: contig of 13240 bp in length  
250123: gap of unknown length  
250222: contig of 6811 bp in length  
250233: gap of unknown length  
257033: gap of unknown length  
257134: contig of 1938 bp in length  
259071: gap of unknown length  
259171: contig of 1837 bp in length  
261008: gap of unknown length  
261108: contig of 3340 bp in length  
264448: gap of unknown length  
264548: contig of 3000 bp in length  
267548: gap of unknown length  
267648: contig of 3834 bp in length  
271482: gap of unknown length  
271582: contig of 4040 bp in length  
275622: gap of unknown length  
275722: contig of 10385 bp in length  
286107: gap of unknown length  
286207: contig of 1911 bp in length.  
286208: contig of 1911 bp in length.

FEATURES  
Location/Qualifiers  
1..288118  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clones="CH230-293111"

misc\_feature  
1..1287  
/note="wgs\_end\_extension  
clone\_end:T7"

misc\_feature  
1338..5778  
/note="wgs\_end\_extension  
clone\_end:T7"

Submitted (11-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Oct 11, 2002 this sequence version replaced gi:21989925.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold,

RESULT 8  
AC121727

LOCUS AC121727 228368 bp DNA linear HTG 26-SEP-2002  
 DEFINITION Rattus norvegicus clone CH230-326G7, \*\*\* SEQUENCING IN PROGRESS  
 AC121727 \*\*\* 13 unordered pieces.  
 AC121727.3 G1:23322328  
 HTG: HTGS PHASE1; HTGS DRAFT; HTGS\_ENRICHED.  
 KEYWORDS Rattus norvegicus (Norway rat)  
 SOURCE Rattus norvegicus  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 REFERENCE 1 (bases 1 to 228368)  
 AUTHORS Muzny, D. Marie, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Angiano, D., Anyalebechi, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Blunk, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyie, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gibrgeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Haves, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorena, L., Loulsged, H., Lozano, R., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwankweli, O., Okunju, G., Olarnpunsagoon, A., Pal, S., Parks, K., Paternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindester, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G., and Gibbs, R.A.  
 Direct Submission  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 228368)  
 AUTHORS Worley, K.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (21-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 REFERENCE 3 (bases 1 to 228368)  
 AUTHORS Rat Genome Sequencing Consortium.  
 TITLE Direct Submission  
 JOURNAL Submitted (26-SEP-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 On Sep 26, 2002 this sequence version replaced gi:21909383.  
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). As a result, the sequence may extend beyond the ends of the clone and there may be contigs that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
 ----- Project Information  
 Center project name: GYGF  
 Center clone name: CH230-326G7  
 ----- Summary Statistics  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 174976 bases at least Q40  
 Consensus quality: 181861 bases at least Q30  
 Consensus quality: 185371 bases at least Q20  
 Estimated insert size: 239944; sum-of-contigs estimation  
 Quality coverage: 3x in Q20 bases; sum-of-contigs estimation

-----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 13 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence.  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 3598: contig of 3598 bp in length  
 \* 3599 3698: gap of unknown length  
 \* 3999 14450: contig of 10752 bp in length  
 \* 14451 14550: gap of unknown length  
 \* 14551 49626: contig of 34976 bp in length  
 \* 49527 49626: gap of unknown length  
 \* 49627 59763: contig of 10137 bp in length  
 \* 59764 59863: gap of unknown length  
 \* 59864 89249: contig of 29386 bp in length  
 \* 89250 89349: gap of unknown length  
 \* 89350 95260: contig of 5911 bp in length  
 \* 95261 95360: gap of unknown length  
 \* 95361 202485: contig of 107125 bp in length  
 \* 202486 202585: gap of unknown length  
 \* 202586 214413: contig of 11827 bp in length  
 \* 214413 214513: gap of unknown length  
 \* 214513 216209: contig of 1697 bp in length  
 \* 216210 216309: gap of unknown length  
 \* 216310 217661: contig of 1452 bp in length  
 \* 217662 217861: gap of unknown length  
 \* 217862 219647: contig of 1786 bp in length  
 \* 219648 219748: gap of unknown length  
 \* 219748 224650: contig of 4902 bp in length  
 \* 224650 224749: gap of unknown length  
 \* 224750 228368: contig of 3619 bp in length.

FEATURES  
 source  
 1. 228368  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-326G7"  
 misc\_feature  
 14551..17292  
 /note="wgs contig"  
 misc\_feature  
 22272..24156  
 /note="wgs contig"  
 misc\_feature  
 40041..43821  
 /note="wgs contig"

misc\_feature 49627..51977  
/note="wgs contig"  
misc\_feature 54058..59763  
/note="wgs contig"  
misc\_feature 89350..90810  
/note="wgs contig"  
misc\_feature 90861..93339  
/note="wgs contig"  
misc\_feature 93390..95260  
/note="wgs contig"  
misc\_feature 95361..97696  
/note="wgs contig"  
misc\_feature 192296..194665  
/note="wgs contig"  
misc\_feature 202586..204894  
/note="wgs contig"  
misc\_feature 212212..214412  
/note="wgs contig"  
BASE COUNT 54283 a 40893 c 40157 g 51976 t 41059 others  
ORIGIN

Query Match 48.3%; Score 842.6; DB 2; Length 228368;  
Best Local Similarity 79.6%; Pred. No. 7.8e-124;  
Matches 1084; Conservative 0; Mismatches 264; Indels 13; Gaps 7;  
Qy 388 TATGATTATGATAAATTTGCCAGGACATCTGTTATCATAGCATTTTATATGAAGCCTGG 447  
Db 163632 TATGATTATGATAAATTTGCCAGGACATCTGTTATCATAGCATTTTATATGAAGCCTGG 163691  
Qy 448 TCAACTCTCTCCGAGCAGTTTACAGTGTCTTTCAGACATCCCGGATATCTCTAGAA 507  
Db 163692 TCCACACTCTTTCAGACAGTTTACAGTGTCTTTCAGACATCCCGGATATCTCTAGAA 163751  
Qy 508 GAAGTGATCTCTTGTAGATGACTACAGTGATAGAGACACCTGAAGAGCGCTTGGCCAAAT 567  
Db 163752 GAGTCAATTTTGTAGATGACTACAGTGATAGAGACACCTGAAGAGCGCTTGGCCAAAT 163810  
Qy 568 GAGCTTTCCGAGCTGCCAAGGTGGCTGTATCGCGCCCAACAGAGAGAGGGCTGGTG 627  
Db 163811 GAGCTGTACAGCTGCCCTAGGTGACCTGTATCCCTGTAGCA--AGAGAGGGCTTAGTG 163868  
Qy 628 CGAGCCGGCTGTGGGGGGCTGTGGCGGAGGGCGATGTTCTGACCTTCTGGACTGT 687  
Db 163869 CAAG-CCGGTACTGGAGAGCTTTGGCGCCAGGGCGAAGTGTCTGGACTGT 163927  
Qy 688 CACTGTGAGTGCCACGAAGGTGGCTGGAGCGCTGTGAGAGGATCCATGAAGAGAG 747  
Db 163928 CACTGTGAGTGCTATGAGGGTGGCTGGAGTCCCTGTGAGAGATCCACGAAGAGAG 163987  
Qy 748 TCGCAGTGTGTGCCCGGTGATTTGATGTATGACACTGGAAACACCTTTGAAATACCTGGG 807  
Db 163988 TCGCCCGTGTGTGCCCGGTATTCATGTATGACTGGAACACCTCTGAGTACCTGGGC 164047  
Qy 808 AACTCCGGGAGGCCAGATCGGGGTTTCAGCTGGAGGCTGGTTTCAGTGCGACACA 867  
Db 164048 AACTCCAGAGAGGCCAGATTTGGTGGCTTTGACTGGCGCTTAGTATTTACGTGGCAGCTG 164107  
Qy 868 GTTCCTGAGAGGAGAGATACGATCAATCCCGCTGATGTGATGCTCAGGTCTCCAAACA 927  
Db 164108 GTCCCTTGGGGAGCGGAGTTGATCGGAGTCTTATGACATTTACAGTCTCCAACT 164167  
Qy 928 ATGGCTGGTGGCTGTTTGTGTGAGTAAGAAATTTTGAATATCTGGGCTCTTATGAT 987  
Db 164168 ATGGCTGGGAGCTGTTTGTGTGAGTAAGAGATTTTGAATATCTGGGCTCTTATGAT 164227  
Qy 988 ACAGGAATGAGTTTGGGAGGAGAAACCTCGAATTTTCTTTAGGATCTGGCAGTGT 1047  
Db 164228 ACCGGAATGAGTCTGGGAGGAGAAACCTTGAATTTCTCTTTAGGATCTGGCAGTGT 164287  
Qy 1048 GGTGGGTTCTGGAACACACACCCATGTTCCCATGTTGGCCATGTTTCCCGCAGCAGCT 1107  
Db 164288 GGT--GACTCTGGAACACACACCCCTTACTCCACGTTGGGCCACGCTTCTCCCTTAAGCAAGCT 164345

Qy 1108 CCTACTCCCGCAACAAGGCTCTGGCCAAACAGTGTTCGTGAGCTGAAGTATGATGGAT 1167  
Db 164346 CCTATTTCAGCAGCAAGGCTCTAGCCAAACAGTGTCCGAGCTGCAGAAAGTGTGACGAT 164405  
Qy 1168 GAATTTAAGAGCTCTACTACCATCGCAACCCCGTCCCGCTTGGAAACCTTTTGGGGAT 1227  
Db 164406 GAATTTAAGAACTCTTACCAACCCCGGAAACCCCG-AGCCCGCTTGGAAACCTCTCTGGGGAG 164464  
Qy 1228 GTGACAGAGAGAGAGAGCTCCGGGCAAGCTCCAGTGTAAAGACTTCAAGTGTCTCTTG 1287  
Db 164465 CTGACAGAGAGAGAGAGCTTTCAGCCCAAGCTCCAGTGTGAAGACTTANAGCGGTTCTG 164524  
Qy 1288 GAGACTGTGTATCCAGAACTGCGATGTCCTGAGACAGGCTTCTTCGGGATGCTC 1347  
Db 164525 GCTACATGTATCCAGAACTGCGATGTCAGAGAGAGAGGCTTGTCTTCTTCGGAAGCTT 164584  
Qy 1348 CAGAACAAAGGACTTACAGACTACTGCTTTTGTACTATTAACCTCCCGATGAACACAGATT 1407  
Db 164585 GAGAACAGAGAGAGAGCGGGGTACTGCTGGATATTAATCCTCCAGTGAAACAGTGTG 164644  
Qy 1408 GTGGGACACCAAGGTCTATTCTGTACTCTGTCTGATGGATGGGCCAGAAATCAGTTTTTCGAG 1467  
Db 164645 GAAGGCCACCAAGGTCTTTTGTACTCTGCGCAGAGATGGTTCAGAACAGTCTTTCGAG 164704  
Qy 1468 TACAGCTCCAGAAAGAAATACGCTATAACACCCACAGCTGAGGCTGCAATGCTGTG 1527  
Db 164705 TGTACATCCCGGCAAGAAATACGCTATAACACCCCGCAGCTCCAGGTCTGCATAGCAGTG 164764  
Qy 1528 GAAGCAGGATGATACCTTTATCATGCTATCTGCGAAGAACTGCCCGCAGAGAAATCAG 1587  
Db 164765 GAGAACAGGAGAGAT--CCTTATTCGGGATCTCTGGAAGATACGGTTCAGAGAAATCAA 164822  
Qy 1588 AAGTTTCATCTTGAGGAGGATGGA---TCTTTATTTACGAAACAGTCCAAGAAATGTGT 1643  
Db 164823 GAGTTTCATCTACAGGAGGAGCAGCAGCTAGTCTCAAGCAGACCGGAAATGTGT 164882  
Qy 1644 CCAGGCTGCGAGGAGAGTGCAGTGCAGTTCGTTTCTTCTTACGAGACTGCACCAA 1703  
Db 164883 GGAGGCCACAGAGAGAGTCAACTCTGCTTTTGTCTCCGAACTTTCGAGACTGTAGCAA 164942  
Qy 1704 CTGGGATCATCAGAAATGTTTCTTCAAGAGCGCATGTTAT 1744  
Db 164943 CTGAGACGCCAGAGGTGTTCTGCCAGGAGCTCATGTCTAT 164983

RESULT 9  
AC098559 7/c  
WPCOMMENT

Sequence split into 8 fragments LOCUS AC098559 Accession AC098559

Fragment Name	Begin	End
AC098559_0	1	110000
AC098559_1	100001	210000
AC098559_2	200001	310000
AC098559_3	300001	410000
AC098559_4	400001	510000
AC098559_5	500001	610000
AC098559_6	600001	710000
AC098559_7	700001	794616

Continuation (8 of 8) of AC098559 from Base 700001 (AC098559 Rattus norvegicus clone CH2)

Query Match 48.2%; Score 841; DB 2; Length 94616;  
Best Local Similarity 79.6%; Pred. No. 1.5e-123;  
Matches 1083; Conservative 0; Mismatches 265; Indels 13; Gaps 7;  
Qy 388 TATGATTATGATAAATTTGCCAGGACATCTGTTATCATAGCATTTTATATGAAGCCTGG 447  
Db 61744 TATGATTATGATAAATTTGCCAGGACATCTGTTATCATAGCATTTTATATGAAGCCTGG 61685  
Qy 448 TCAACTCTCTCCGAGCAGTTTACAGTGTCTTTCAGACATCCCGGATATCTCTGCTAGAA 507  
Db 61684 TCCACTCTCTTCAGACAGTTTACAGTGTCTGAGACTTCCCGCTGATATCTCTGCTGAG 61625  
Qy 508 GAAGTGATCTCTGTAGATGACTACAGTGATAGAGACCTGAAGGAGCGCTTGGCCAAT 567



TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

Weinstein, G. and Gibbs, R.A.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 288118)  
 Worley, K.C.  
 Direct Submission  
 Submitted (27-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 288118)  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (11-OCT-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Oct 11, 2002 this sequence version replaced gi:21998925.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described  
 in the feature table below represents a scaffold in the Atlas  
 assembly (a 'contig-scaffold'). Within each contig-scaffold,  
 individual sequence contigs are ordered and oriented, and separated  
 by sized gaps filled with Ns to the estimated size. The sequence  
 may extend beyond the ends of the clone and there may be sequence  
 contigs within a contig-scaffold that consist entirely of whole  
 genome shotgun sequence reads. Both end sequences and whole genome  
 shotgun sequence only contigs will be indicated in the feature  
 table.

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: http://www.hgsc.bcm.tmc.edu/  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GRGM  
 Center clone name: CH230-293111  
 ----- Summary Statistics  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 214989 bases at least Q40  
 Consensus quality: 222817 bases at least Q30  
 Consensus quality: 227320 bases at least Q20  
 Estimated insert size: 252471; sum-of-contigs estimation  
 Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

-----  
 \* NOTE: Estimated insert size may differ from sequence length  
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html)  
 \* NOTE: This sequence may represent more than one clone.  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 14 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1  
 5778: contig of 5778 bp in length  
 5879: gap of unknown length  
 16373: contig of 10495 bp in length  
 16473: gap of unknown length  
 16474: gap of unknown length  
 90421: contig of 73947 bp in length  
 90520: gap of unknown length  
 236783: contig of 146262 bp in length  
 236882: gap of unknown length  
 250122: contig of 13240 bp in length  
 250123: gap of unknown length  
 250223: contig of 6811 bp in length  
 257034: gap of unknown length  
 257133: contig of 1938 bp in length  
 259072: gap of unknown length  
 259171: contig of 1837 bp in length  
 261009: gap of unknown length  
 261108: gap of unknown length  
 264448: contig of 3340 bp in length  
 264449: gap of unknown length

264549 267548: contig of 3000 bp in length  
 267549 267648: gap of unknown length  
 267649 271482: contig of 3834 bp in length  
 271483 271582: gap of unknown length  
 271583 275622: contig of 4040 bp in length  
 275623 275722: gap of unknown length  
 275723 286107: contig of 10385 bp in length  
 286108 286207: gap of unknown length  
 286208 288118: contig of 1911 bp in length.  
 FEATURES  
 Location/Qualifiers  
 source  
 1..288118  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-293111"  
 misc\_feature  
 1..1287  
 /notes="wgs end extension  
 clone end:T7"  
 misc\_feature  
 1338..5778  
 /notes="wgs end extension  
 clone end:T7"  
 misc\_feature  
 3086..3988  
 /notes="clone boundary  
 clone end:T7"  
 misc\_feature  
 19118..21281  
 /notes="wgs contig"  
 misc\_feature  
 5879..7959  
 end sequence:RXAIQ54TJB"  
 /notes="wgs contig"  
 misc\_feature  
 12280..13622  
 /notes="wgs contig"  
 misc\_feature  
 14019..16373  
 /notes="wgs contig"  
 misc\_feature  
 16474..19067  
 /notes="wgs contig"  
 misc\_feature  
 19118..21281  
 /notes="wgs contig"  
 misc\_feature  
 21332..22964  
 /notes="wgs contig"  
 misc\_feature  
 28345..31644  
 /notes="wgs contig"  
 misc\_feature  
 31695..34111  
 /notes="wgs contig"  
 misc\_feature  
 40505..44913  
 /notes="wgs contig"  
 misc\_feature  
 44964..47869  
 /notes="wgs contig"  
 misc\_feature  
 47920..49820  
 /notes="wgs contig"  
 misc\_feature  
 49871..52672  
 /notes="wgs contig"  
 misc\_feature  
 52723..56843  
 /notes="wgs contig"  
 misc\_feature  
 61887..63717  
 /notes="wgs contig"  
 misc\_feature  
 72165..76565  
 /notes="wgs contig"  
 misc\_feature  
 78769..86319  
 /notes="wgs contig"  
 misc\_feature  
 86370..90420  
 /notes="wgs contig"  
 misc\_feature  
 90521..92438  
 /notes="wgs contig"  
 misc\_feature  
 92489..94959  
 /notes="wgs contig"  
 misc\_feature  
 96097..99542  
 /notes="wgs contig"  
 misc\_feature  
 236883..239542

Query Match 48.2%; Score 841; DB 2; Length 288118;  
 Best Local Similarity 79.6%; Pred. No. 1.4e-123;  
 Matches 1083; Conservative 0; Mismatches 265; Indels 13; Gaps 7;  
 QY 388 TATGATTATGATAATTGGCCAGGACATCTGTTATCATAGCATTTTATATGAAGCCTGG 447







ERKLRDCKLOCFKPELETVYVPELVHVPEDRPGFGLMQLNKLTDYCFDYNPPDENOI  
VGHVILYLCHMGCONQFPEVTSQKEIRYNTHQPEGCIAREGMDTLIMHLCEBTABE  
NQKFILEDGSLFHEQKCKVQARKESSDSFVPLLRDCTNSDHQKWFPEKRMIL"  
BASE COUNT 578 a 372 c 433 g 545 t  
ORIGIN

Query Match 47.5%; Score 830; DB 9; Length 1928;  
Best Local Similarity 99.4%; Pred. No. 1.3e-121;  
Matches 833; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
909 TGTATCATCAGGCTCCCAACAATGGCTGGTGGCTGTTGCTGTGAGTAAGAATAATTTGA 968  
Db TGCCATGTGGTCTCCACAATGGCTGGTGGCTGTTGCTGTGAGTAAGAATAATTTGA 133  
969 ATATCTGGGCTTTATGATACAGGAATGGAAGTTTGGGAGGAGAGAAACCTCGAATTTTC 1028  
Db ATATCTGGGCTTTATGATACAGGAATGGAAGTTTGGGAGGAGAGAAACCTCGAATTTTC 193  
1029 CTTTATGAGATCTGGCAGTGTGGTGGTCTTGGAACAACACCCATGTTCCCATGTTGGCCA 1088  
Db CTTTATGAGATCTGGCAGTGTGGTGGTCTTGGAACAACACCCATGTTCCCATGTTGGCCA 253  
1089 TGTGTTTCCCAAGCAAGCTCCCTACTCCCGCAACAAGGCTCTGGCCAACAGTGTTCGTGC 1148  
Db TGTGTTTCCCAAGCAAGCTCCCTACTCCCGCAACAAGGCTCTGGCCAACAGTGTTCGTGC 313  
1149 AGCTGAAGTATGAGTGAATTTAAAGAGCTCTACTACCATCGCAACCCCGTGCCTG 1208  
Db AGCTGAAGTATGAGTGAATTTAAAGAGCTCTACTACCATCGCAACCCCGTGCCTG 373  
1209 CTTGGACCTTTTGGGATGTGACAGAGAGAAAGAGCTCCGGCAACAAGTCTCCAGTGTAA 1268  
Db CTTGGACCTTTTGGGATGTGACAGAGAGAAAGAGCTCCGGCAACAAGTCTCCAGTGTAA 433  
1269 AGACTTCAAGTGGTCTTGGAGAGTGTGTATCCAGAACTGTGCTGCTGAGGACAGGCC 1328  
Db AGACTTCAAGTGGTCTTGGAGAGTGTGTATCCAGAACTGTGCTGCTGAGGACAGGCC 493  
1329 TGGCTTCTTGGGATGTCCAGAAACAAGGACTAACAGACTACTGCTTTGACTATAACCC 1388  
Db TGGCTTCTTGGGATGTCCAGAAACAAGGACTAACAGACTACTGCTTTGACTATAACCC 553  
1389 TCCGATGAAACACAGATTTGGGACACCAAGTCTATTTGTACCTCTGTATGGATGGG 1448  
Db TCCGATGAAACACAGATTTGGGACACCAAGTCTATTTGTACCTCTGTATGGATGGG 613  
1449 CCAGAAATCAGTTTTCAGTACAGCTCCGAAAGAAATACGCTATAACACCCACAGCC 1508  
Db CCAGAAATCAGTTTTCAGTACAGCTCCGAAAGAAATACGCTATAACACCCACAGCC 673  
1509 TGAGGGTGCATTTCTGGAAGCAGGAATGGATACCTTTATCATGCTCTCTGCGAAGA 1568  
Db TGAGGGTGCATTTCTGGAAGCAGGAATGGATACCTTTATCATGCTCTCTGCGAAGA 733  
1569 AACTGCCCCAGAGATCAGAAATTCATCTTGCAGAGATGGATCTTTATTTACGAACA 1628  
Db AACTGCCCCAGAGATCAGAAATTCATCTTGCAGAGATGGATCTTTATTTACGAACA 793  
1629 GTCCAGAAATGTCTCAGAGCTCGAGAGAGGAGTCCAGTGACAGTTTGGTTCACCTTT 1688  
Db GTCCAGAAATGTCTCAGAGCTCGAGAGAGGAGTCCAGTGACAGTTTGGTTCACCTTT 853  
1689 ACGAGACTGCACCAACTCCGATCATCAGAAATGTTCTTCAAAGAGGCGCATGTTATGA 1746  
Db ACGAGACTGCACCAACTCCGATCATCAGAAATGTTCTTCAAAGAGGCGCATGTTATGA 911

RESULT 12

AC099237\_3

WPCOMMENT

Sequence split into 6 fragments LOCUS AC099237 Accession AC099237

Fragment Name

Begin

End

AC099237\_0

1 110000

AC099237\_1 100001 210000  
AC099237\_2 200001 310000  
AC099237\_3 300001 410000  
AC099237\_4 400001 510000  
AC099237\_5 500001 545260  
Continuation (4 of 6) of AC099237 from base 300001 (AC099237 Rattus norvegicus clone CH2:

Query Match 46.5%; Score 812.6; DB 2; Length 110000;  
Best Local Similarity 79.2%; Pred. No. 4.7e-119;  
Matches 1077; Conservative 0; Mismatches 269; Indels 13; Gaps 9;  
388 TATGATTTATCATATTTGCCCAGGACATCTGTTATCATAGCATTTTATATGAGCCTGG 447  
Db TATGATTTATGATTAACCTGCCCCAAGACATCCGTTGTCTAGCATTTTATATGAGCCTGG 43652  
448 TCAACTCTCTCTCGGACAGTTTACAGTGTCTTGGAGACATCCCGGATATCTCTCTAGAA 507  
Db TCCACACTCTCTTACAGAGTTTACAGTGTCTCGAGACTTCCCTGTATATCTCTGTGAG 43712  
508 GAAGTGTATCTTGTAGATGACTACAGTGTAGAGAGCACCTGAAGAGCGCTTGGCCAAT 567  
Db GAGTCAATCTGTGTAGATGACTACAGTGTAGAGAGTACCTGTACCTGCTGTAGCA - 43771  
568 GAGCTTTCGGGACTGCCAAGGTGCGCTGATCCGCGCCCAACAGAGAGAGGCGCTGGTG 627  
Db GAGCTGTACAGCTGCGCTAGGTGTACCTGTATCCCTGTGTAGCA - AGAGAGGCGCTAGTG 43829  
628 CGAGCCCGGTGTGGGGGGCTCTGGCGGAGGGGCGATGTTCTGACCTTCTCTGACTGT 687  
Db CAAG - CGGGTACTTGGAGGCTCTGGGGCCAGGGGCGAAGTGTGAGGTTTCTGGACTGT 43888  
688 CACTGTGAGTGCCACGAAGGGTGGTGGAGCCGCTGCTGCAGAGGATCCATGAAGAGGAG 747  
Db CACTGTGACTGTCTAGAGGGTGGCTGGAG - TCTGCTGCAGAGGATCCACGAGAGGAG 43946  
748 TCGCAGTGTGTCCCGGTGATGATGATCGATCGATGGAACACCTTCGAAATACCTGGGG 807  
Db TCGCCGCTGTGTGCGCGGTATTCGATGCTGACTGGAACACTTCTGAGTACTCTGGGC 44006  
808 AACTCCGGGAGGCCACAGATCGCGGTTCGACTGGAGGCTGGTGTTCAGTGCGACACA 867  
Db AACTCC - GGAAGCCCAAGATTGAGGCTTTGACTGGCGCTTAGTAITTCAGTGGCACGTG 44065  
868 GTTCTTGAGAGGAGGAGGATACGAGATGCAATCCCGCTCGATGTCTATCAGGTCTCCAA 927  
Db GTCCCTTGGCGGAGCGGAAGTTGATGCGGACTCTTATTCACATATCAGGTCTCCAACT 44125  
928 ATGCTGTGGGTGTTGTTGCTGTGATGAGAAATATTTGAAATATCTGGGGTCTTATGAT 987  
Db ATGCTGTGGGACTGTTGTTGTTGATGAGAGATATTTTGAATACCTGGGGTCTTATGAT 44185  
988 ACAGGAATGGAAGTTTGGGAGGAGAAACCTCGAAATTTTCTTTAGGATCTGGCAGTGT 1047  
Db ACCGGAATGGAAGTTAGGGAGGAGAAACCTTGAATTTCTCTTTAGGATCTGGCAATGT 44245  
1048 GGTGGGCTTCGAAACACACACCTCATGTTCCCATGTTGGCCATGTTTCCCAAGCAAGCT 1107  
Db GGTGACTCTCTGAAACACACACCTTACTCCACGTTGGCCACCGTCTTCCCTTAGCAAGCT 44305  
1108 CCTACTCTCCGCAACAGGCTCTGGCCCAACAGTGTGTCGACGTGAAGTATGATGAT 1167  
Db CCTATTTACGCAAGAGGCTCTAGCCAAACAGTGTCCGAGCTGCAGAAAGTGTGGCGGAT 44365  
1168 GAATTTAAAGAGCTTACTACCATCCCAACCCCGTCCCGCTTGGACCTTTTGGGAT 1227  
Db GAATTTAAAGAGCTTACTACCATCCCAACCCCGTCCCGCTTGGACCTTTTGGGAT 44366  
1228 GTGACAGAGAGGAGGAGCTCCCGGCAACAGCTCCAGTGTAAAGACTTCAAGTGTCTTG 1287  
Db GTGACAGAGAGAGGAGGCTTCCAGCCACAGCTCCAGTGTGAAGATTTAAAGCGGCTCTG 44484  
1288 GAGACTGTGTATCCAGAACTGCTGATGCTGAGGACAGGCTGCTTCTTGGGATGCTC 1347  
Db GAGACTGTGTATCCAGAACTGCTGATGCTGAGGACAGGCTGCTTCTTGGGATGCTC 1347

```

Db 44485 GCTACACTGTACCCAGAACTGCATGTGCCAGAGACAGGCGCTGTTCTTCGGGAAGCTT 44544
Qy 1348 CAGAACAAAGACTAAACAGACTACTGCTTTGACTATATACCTCCCGATGAACACAGATT 1407
Db 44545 GAGAACAGAGACTACGGGGTACTGCTGACTATATATCTCCAGTGAACAACTGTG 44604
Qy 1408 GTGGGACACCAAGGTCATTCTGTACTCTCTGTCATGGGATGGGCCAAGATCAAGTTTTCGAG 1467
Db 44605 GAAGGCCACCAAGGTCCTTTTGTACTCTGTCACGGGATGGTTCAGAACCAAGTTTTCGAG 44664
Qy 1468 TACACGTCCTCCAGAAAGAAATACGCTATAACACCCACAGCCTGAGGGCTGATTCGCTGTG 1527
Db 44665 TGTACATCCCAAGCAAGAAATACGCTATAACCA-TGCCAGCCCGCCAGGCTGTCATACGACAGT 44723
Qy 1528 GAACAGCAATGGATACCTTATCATGCTCTGCGAAGAACTGCCCGACAGAAATCGAG 1587
Db 44724 GAGGAAGGGAAGGAT--CCCTATCCCGGCTCTCGGAAGATACCGTTCCAGAGAACTCA 44781
Qy 1588 AAGTTTCATCTTGCAGGAGGATGATCTTTATTTTCCAGCAAGTCCCAAGAAATGTGTCAG 1647
Db 44782 GAGTTTCATCTTGCAGGAGGACCGACCTTAGTCTCAAGCAGAGCAGGAAATGTGTCAG 44841
Qy 1648 GCTCGAGGAGGAGTGCAGTGACAGTTTCCTTCCATCTTTACAGACTGACCAACTCG 1707
Db 44842 GCCACAGAGGAGGAGTTCACCTGCTGCTTCCGAACTTGCAGACTGTAGCAACTCA 44901
Qy 1708 G--ATCATCAAAATGGTTCTTCAAGAGCCGATGTTAT 1744
Db 44902 GACACAGCCAGAGGTGGTCTCAAGGAGTGCATGTGCAT 44940

```

## RESULT 13

```

AC134060/c
LOCUS AC134060 166547 bp DNA linear HTG 22-SEP-2002
DEFINITION Rattus norvegicus clone CH230-227M12, *** SEQUENCING IN PROGRESS
***, 33 unordered pieces.
ACCESSION AC134060
VERSION AC134060.1 GI:23267331
KEYWORDS HTG; HTGS_PHASE1. (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

```

## REFERENCE

1 (bases 1 to 166547)

Murphy, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Albrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biwalo, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, I., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, J., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idiebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpachy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshuwa, L., Louisedge, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mayhew, S., McLeod, M., McNeill, T., Meenen, E., Milosavljevic, A.,

Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokemele, O., Okwuonu, G., Olarunpunsagoon, A., Pal, S., Parke, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Reigh, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajda, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorrelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, P., Williams, G., Willson, R., Wleccyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Unpublished

2 (bases 1 to 166547)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (22-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

----- Project Information

Center project name: GTSQ

Center clone name: CH230-227M12

----- Summary Statistics

Sequencing vector: Plasmid;

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 142428 bases at least Q40

Consensus quality: 149083 bases at least Q30

Consensus quality: 153531 bases at least Q20

-----

\* NOTE: Estimated insert size may differ from sequence length

\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 33 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

-----

\* 1215: contig of 1215 bp in length

\* 1315: gap of unknown length

\* 1316: contig of 1512 bp in length

\* 2828: gap of unknown length

\* 2928: contig of 1237 bp in length

\* 4165: gap of unknown length

\* 4264: contig of 1563 bp in length

\* 5828: gap of unknown length

\* 5928: contig of 2730 bp in length

\* 8658: gap of unknown length

\* 8757: contig of 1208 bp in length

\* 9965: gap of unknown length

\* 10065: contig of 2717 bp in length

\* 12782: gap of unknown length

\* 12883: contig of 1702 bp in length

\* 14585: gap of unknown length

\* 17098: contig of 2414 bp in length

\* 17198: gap of unknown length

\* 17099: contig of 1279 bp in length

\* 17199: contig of 1279 bp in length

```
* 18478 18577: gap of unknown length
* 18578 21168: contig of 2591 bp in length
* 21169 21268: gap of unknown length
* 21269 23708: contig of 2440 bp in length
* 23709 23808: gap of unknown length
* 23809 26530: contig of 2722 bp in length
* 26531 26630: gap of unknown length
* 26631 28801: contig of 2171 bp in length
* 28802 28901: gap of unknown length
* 28902 33091: contig of 4190 bp in length
* 33092 33191: gap of unknown length
* 33192 36968: contig of 3777 bp in length
* 36969 37068: gap of unknown length
* 37069 39304: contig of 2236 bp in length
* 39305 39404: gap of unknown length
* 39405 43433: contig of 4029 bp in length
* 43434 43534: gap of unknown length
* 43534 46236: contig of 2703 bp in length
* 46237 46336: gap of unknown length
* 46337 50699: contig of 4363 bp in length
* 50700 50799: gap of unknown length
* 50800 53947: contig of 3148 bp in length
* 53948 54047: gap of unknown length
* 54048 58293: contig of 4246 bp in length
* 58294 58393: gap of unknown length
* 58394 63160: contig of 4767 bp in length
* 63161 63260: gap of unknown length
* 63261 68334: contig of 5074 bp in length
* 68335 71462: contig of 5028 bp in length
* 71463 73562: gap of unknown length
* 73563 79617: contig of 6055 bp in length
* 79618 86568: contig of 6851 bp in length
* 86569 86668: gap of unknown length
* 86669 95652: contig of 8984 bp in length
* 95653 95753: gap of unknown length
* 95753 105300: contig of 9548 bp in length
* 105301 112938: contig of 7538 bp in length
* 112939 113038: gap of unknown length
* 113039 122447: contig of 9409 bp in length
* 122448 122548: gap of unknown length
* 122549 137894: contig of 15347 bp in length
* 137895 137995: gap of unknown length
* 137995 166547: contig of 28553 bp in length.

FEATURES             Location/Qualifiers
     source            1..166547
     /organism="Rattus norvegicus"
     /mol_type="genomic DNA"
     /db_xref="taxon:10116"
     /clone="CH230-227M12"

BASE COUNT  45529 a 35358 c 35370 g 45112 t 5178 others

Query Match      46.5%; Score 812.6; DB 2; Length 166547;
Best Local Similarity 79.2%; Pred. No. 4.5e-119;
Matches 1077; Conservative 0; Mismatches 269; Indels 13; Gaps 9;

Qy 388  TATGATTATGATTAATTTGCCAGAGACATCTGTTATCATAGCATTTTATATGAAGCCTGG 447
      |||
Db 84617 TATGATTATGATTAATTTGCCAGAGACATCTGTTATCATAGCATTTTATATGAAGCCTGG 84558

Qy 448  TCAACTCTCTTCCGAGAGTTTACAGTGTCTTTCAGACATCCCGGATATCCTGCTAGAA 507
      |||
Db 84557 TCCACTCTCTTCCGAGAGTTTACAGTGTCTTTCAGACATCCCGGATATCCTGCTGAG 84498

Qy 508  GAAGTGATCTCTTGTAGATGACTACAGTGATAGAGACACCTGAAGAGCGCTTGGCCAAAT 567
      |||
Db 84497 GAGTGATCTCTTGTAGATGACTACAGTGATAGAGACACCTGAAGAGCGCTTGGCCAAAT 84439

Qy 568  GAGCTTTGGGAGCTGCGCAAGGTGCGCTGATCCGCGCCCAACAAGAGAGAGGCGCTGGTG 627
      |||
```

```
Db 84438 GAGCTGTACAGCTGCCCTAGGTGTACCTGATCCCTGGTAGCA--AGAGAGGGGCTTAGTG 84381
Qy 628  CGAGCCCGGCTGTGGGGGGGCTCTCGCGCGAGGGGGCGATGTTCTGACCTTCTCGACTGT 687
      |||
Db 84380 CAAG-CCGGCTACTGGGAGCCTCTGGGGCCAGGGGCGAAGTGTGCTGAGGTTTCTGACTGT 84322
Qy 688  CACTGTGAGTGCACAGGAAGGGTGGCTGGAGCCGCTGCTGCAGAGGATCCATGAAGAGGAG 747
      |||
Db 84321 CACTGTGACTGTCTATGAGGGGTGGCTGGAG--TCTGCTGCAGAGGATCCAGAGAGGAG 84264
Qy 748  TCGGCACTGTGTGCCCGGTGATTGATGTGATCGACTGGAGCAACCTTTCGAATACCTCGGG 807
      |||
Db 84263 TCGGCGGTGTGTGCCCGGTATCGATGTCAATTGACTGGAACTTCTGAGTCACTGGGC 84204
Qy 808  AACTCCGGGAGCCCCAGATCGGCGGTTTCGACTGGAGGCTGTGTTCAGCTGTCACACACA 867
      |||
Db 84203 AACTCC--GGAAGCCCCAGATTGGAGGCTTTGACTGGCGCCTAGTATTACGTGSCACGTG 84145
Qy 868  GTTCTGAGAGGAGGAGATACGGATGCAATCCCCCGCTCGATGTCATCAGGTCTCCAAACA 927
      |||
Db 84144 GTCCCTTTGGCGGGAGCGGAAGTTGATGCGGACTCTATTGACATTTATCAGGTCTCCAAC 84085
Qy 928  ATGCTGTGTGGGCTGTTTGTGCTGTGAGTAAGAAATATTTTGAATATCTGGGGTCTTATGAT 987
      |||
Db 84084 ATGCTGTGGGACTGTTTGTGTGAGTAAGAGATATTTTGAATACCTGGGGTCTTATGAT 84025
Qy 988  ACAGGAATGGAAGTTTGGGAGGAGAAAAACCTCGAATTTTCTTTAGGATCTGGCAGTGT 1047
      |||
Db 84024 ACCGGAATGGAAGTCTAGGAGGAGAAAAACCTTGAATTTCTTTTAGGATCTGGCAATGT 83965
Qy 1048 GGTGGGGTTCGTGAAACACACACCATGTTCCCATGTTGGCCATGTTTTCCTCCCAAGCAAGCT 1107
      |||
Db 83964 GGTGACTCTCTGGAACACACACCCTTACTCCACAGTGGGCGCACGCTTCTCCCTAAGCAAGCT 83905
Qy 1108 CCTACTCCCGCAACAGGCTCTGGCCCAACAGTGTTCGTGCAGCTGAAGTATGATGAT 1167
      |||
Db 83904 CCCTATTACGACAGCAAGGCTCTAGCCCAACAGTGTCCGAGCTGCAGAAAGTGTGGCGGAT 83845
Qy 1168 GAATTTAAAGAGCTCTACTACCATCGCAACCCCGCTCGCCGCTTGGAACTTTTGGGGAT 1227
      |||
Db 83844 GAATTTAAAGAGCTCTACCAACCCCGGAAATCCCC-AGCCCGCTTGGAACTTCTGGGAG 83786
Qy 1228 GTGACAGAGAGGAGAGCTCCCGGCAACAGCTCCAGGTGTAAAGACTTCAAGTGTGTTCTTG 1287
      |||
Db 83785 CTGACAGAGAGAAAGAGCTTTCAGCCCAAGCTCCAGTGTGAAGATTTAAAGCGGTTCTCTG 83726
Qy 1288 GAGACTGTGTATCCAGAACTGCGATGTCCTGAGAGCAGGCTGCTTCTTCGGGATGCTC 1347
      |||
Db 83725 GCTACACTGTACCCAGAACTGCGATGTGCCAGAGGACAGGCGCTGTTCTTCGGGAAAGCTT 83666
Qy 1348 CAGAACAAAGGACTTAAACAGACTACTGCTTTGACTATAAACCTCCCGATGAAACCAAGATT 1407
      |||
Db 83665 GAGAACAGAGGACTACGGGGGTACTGCTGAGCTATATCTCCAGTGAAACAAATGTG 83606
Qy 1408 GTGGGACACAGGTCATTTCTGTACCTCTGTGATGGGATGGGCCAGAAATCAGTTTTCGAG 1467
      |||
Db 83605 GAAGGCCACCAGTCTCTTTTGTACCTCTGCGACCGGATGGGTGAGAACCAAGTTTTCGAG 83546
Qy 1468 TACAGCTCCAGAGAAAGAAATACGCTATACACCCCAACAGCCTGAGGCTGATGCTGTG 1527
      |||
Db 83545 TGTACATCTCCAGCAAGAAATACGGTATAACA-TCGCCAGGCCAGGCTCTCATAGCAGTG 83487
Qy 1528 GAAGCAGGAATGATACCTTTATCATGCACTCTCGCGAAGAACTGCCCCAGAGAAATCAG 1587
      |||
Db 83486 GAGGAAGGGAAGAT--CCTTATCCCGGGTCTCTGGAAGATACCGTTCAGAGAAATCAA 83429
Qy 1588 AAGTTCATCTTGCAGGAGGATGATCTTTATTTTACAGAAACAGTCCAAAGAAATGTGTCAG 1647
      |||
Db 83428 GAGTTCATCTCAGGAGGACCGCACTTAGTCTCAACAGCAGAGCAGGAATGTGTGAG 83369
Qy 1648 CTTGCGAGGAGAGGAGTGCAGTGAAGTTTGTGTTTCCACTCTTTACAGACTGCAACAACTCG 1707
      |||
Db 83368 GCCACAGAGGAGGAGTTCAACTCTGCTTTGCTCCGAACTTGGGAGACTGTAGCAACTCA 83309
```



\* 58394 63160: contig of 4767 bp in length  
\* 63161 63260: gap of unknown length  
\* 68334 68334: contig of 5074 bp in length  
\* 68435 68435: gap of unknown length  
\* 73462 73462: contig of 5028 bp in length  
\* 73463 73463: gap of unknown length  
\* 73563 73563: contig of 6055 bp in length  
\* 79618 79618: gap of unknown length  
\* 79718 79718: contig of 6851 bp in length  
\* 86569 86569: gap of unknown length  
\* 86668 86668: contig of 8984 bp in length  
\* 95652 95652: gap of unknown length  
\* 95653 95653: contig of 9548 bp in length  
\* 105300 105300: gap of unknown length  
\* 105401 105401: contig of 7538 bp in length  
\* 112939 112939: gap of unknown length  
\* 113039 113039: contig of 9409 bp in length  
\* 122448 122448: gap of unknown length  
\* 122548 122548: contig of 15347 bp in length  
\* 137895 137895: gap of unknown length  
\* 137995 137995: contig of 28553 bp in length.

## FEATURES

source

1. 166547  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clones="CH230-227M12"

BASE COUNT 45529 a 35358 c 35370 g 45112 t 5178 others  
ORIGIN

Query Match 46.1%; Score 805.4; DB 2; Length 166547;

Best Local Similarity 78.8%; Pred. No. 6.1e-118;  
Matches 1073; Conservative 0; Mismatches 273; Indels 15; Gaps 9;

Qy 388 TATGATTATGATATTTGGCCAGGACATCTGTTATCATAGCATTTTATATGAAGCCTGG 447  
Db TATGATTATGATATTTGGCCAGGACATCTGTTATCATAGCATTTTATATGAAGCCTGG 99957  
Qy 448 TCAACTCTCTTCGGACAGTTTACAGTGTCTTGAGACATCCCGGATATCTCTAGAA 507  
Db TCCACACTCTCTTCAGACAG-TTACAGTGTCTCGAGACTTCCCTCTGATATCTCTCGAG 100016  
Qy 508 GAAGTGATCTTTGATGATGACTACAGTGATAGAGACACCTGAAGAGCGCTTGCCCAAT 567  
Db TATGATCTTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 100017  
Qy 568 GAGCTTTCCGAGCTGCGCCAGGTCGCTGATCCGCGCAACAGAGAGAGGGCTCTGCTG 627  
Db GAGCTGTACAGCTGCCNTAGGTGCACTGATCTCTGTAGCA--AGAGAGGGCTTAGT 100133  
Qy 628 CGAGCCCGCTGCTGGGGGCGTCTGCGCGAGGGGCGATGTTCTGACCTTCTCTGGACTGT 687  
Db CAAG-CCGGCTACTGGGAGCCTTTGCGGCCAGGGGCGAAGTGTCTGAGTTTCTGACTGT 100192  
Qy 688 CACTGTGAGTCCACGAAGGTGCTGGAGCGCTGCTGCGAGAGGATCCATGAAGAGGAG 747  
Db CACTGTGACTGTCTATGAGGGTGGCTGGAGTCCCTGCTGCGAGAGATCCAGAGAGGAG 100252  
Qy 748 TCGGCACTGTGTGCCCGGTGATTGATGTGATCGACTGGAACACCTTCGAATACCTCGGG 807  
Db TCGGCGGTGTGTGCCCGGTGATTGATGTGATCGACTGGAACACCTTCGAATACCTCGGG 100312  
Qy 808 AACTCCGGGAGCCCCAGATCGCGGTTTTCAGCTGGAGGCTGGTGTTCAGCTGGCACACA 867  
Db AACTCCAGAGAGCCCCAGATTGGTGGCTTTGACTGGCGCTAGTATTACAGTGGCACGTG 100372  
Qy 868 GTTCTGAGAGGGAGAGATACGAGTGAATCCCGCTCGATGTCATCAGGTCTCCAAACA 927  
Db GTTCTGAGAGGGAGAGAGTGAATCCCGCTCGATGTCATCAGGTCTCCAAACT 100432  
Qy 928 ATGCTCGTGGCTGTTTCTGCTGAGTGAAGAAATATTTGAAATATCTGGGCTCTATGAT 987  
Db ATGCTCGTGGGAGCTGTTTGTGAGTGAAGAGATATTTGAAATACCTGGGCTCTATGAT 100492

Qy 988 ACAGGAATGAAGTTTGGGAGGAGAAAACTCGAAATTTCTTTTAGGATCTGCACTGT 1047  
Db ACCGGAATGAAGTCTGGGAGGAGAAAACTCGAAATTTCTTTTAGGATCTGCACTGT 100552  
Qy 1048 GGTGGGTTCTGGAACACACACCCTATGTTCCCATGTTGGCCATGTTTTCCTCCCAAGCAAGCT 1107  
Db GGT--GACTCAGGAACACACACCCTTACTCCCAAGTGGGCTATTTCTTCCCTTAAGCAAGCT 100610  
Qy 1108 CCTACTCTCCGCAACAGGCTCTGGCCAAACAGTGTTCGTGCAGCTGAAGTATGATGAT 1167  
Db CCCTATTTCACGCAAGGCTCTAGCCAAACAGTGTCCGAGCTGCAGAAAGTGTGACCGAT 100670  
Qy 1168 GAATTTAAAGAGCTCTACTACCATCGCAACCCCGTGCCTTTGGAAACCTTTTGGGAT 1227  
Db GAACTTTAAAGAGCTCTACTACCAACCCCGGAAACCCCG-AGCCCGCTCGGAACCTCTTGGGAG 100729  
Qy 1228 GTGACAGAGAGGAGAGCTCCGGGCAACAGCTCCAGGTGAAGACTTCAAGTGGTCTTGT 1287  
Db CTGACAGAGAGAAAGAGCTTCAGCCCAAGCTCCAGTGTGAAGACTTAAAGCGGTTCTGT 100789  
Qy 1288 GAGACTGTGTATCCAGAACTGTCATGTGCTGAGGACAGGCTTGCTTCTTCGGGATGCTC 1347  
Db GCTACNATGTACCAGAACTGTCATGTGCGAAGAGACAGGCTGGTTCCTTCGGGAACTT 100849  
Qy 1348 CAGAACAAAGAGCTAAACAGACTACTGTTTGACTATAACCTCCCGATGAAGAACCAAGAT 1407  
Db GAGAACAGAGAGAGCGAGGGGGTACTGGCTGGAATATAATCTCCAGTGAACACAGTGTGG 100909  
Qy 1408 GTGGACACCAGGCTCATTTCTGTACTCTGTGTCATGGGATGGCCAGGATCAGTTTTCGAG 1467  
Db AAGGGACCCAGGTC-TTTTGTACTCTGCGCAGAGATGGGTGAGAACCAAGTTCCTTCGAG 100968  
Qy 1468 TACACGTCCCAAGAAATACGCTATAACACCCAGCCTGAGGGCTGCAATGCTGTG 1527  
Db TGTACATCCCGCAGAAATACGCTATAACACCCAGCCTGAGGGCTGCAATGCTGTG 101028  
Qy 1528 GAAGAGGAATGATACCTTTATCATGCTCTCTGCGAAGAAACTGCCCCAGAGAAATCAG 1587  
Db GAGGAGGGAAGAT--CCTTATTCCGATCTCTGGAAGATACGGTTCAGAGAAATCAA 101086  
Qy 1588 AAGTTCATCTTGAGGAGATGA---TCTTTATTTCAGAACAGTCCAGAAATGCT 1643  
Db GAGTTCATCTTACAGGAGGAGCAGACAGCTTAGTCTACAAGCAGACCGGAAATGCT 101146  
Qy 1644 CCAGGCTGCGAGGAGGAGTGCAGTGTGACAGTTCCTGCTTCCACTCTTACAGACTGCACAA 1703  
Db GGAGGCCACAGAGAGAGTTCAACTGCTGCTTTCCTCCAACTTGCAGACTGTAGCAA 101206  
Qy 1704 CTCGGATCATCAGAAATGTTCTTCAAGAGCGCATGTTAT 1744  
Db CTCAGACAGCCAGAGTGGTTCCTCCAGGAGCTCATGTCAT 101247

## RESULT 15

AC098559.1/c

## WPCOMMENT

Sequence split into 8 fragments LOCUS AC098559 Accession AC098559

Fragment Name	Begin	End
AC098559_0	1	110000
AC098559_1	100001	210000
AC098559_2	200001	310000
AC098559_3	300001	410000
AC098559_4	400001	510000
AC098559_5	500001	610000
AC098559_6	600001	710000
AC098559_7	700001	794616

Continuation (2 of 8) of AC098559 from base 100001 (AC098559 Rattus norvegicus clone CH2:

Query Match 46.0%; Score 803.8; DB 2; Length 110000;

Best Local Similarity 78.4%; Pred. No. 1.2e-117;

Matches 1065; Conservative 0; Mismatches 282; Indels 12; Gaps 8;

```
Qy 388 TATGATTATGATAAATTTGGCCAGGACATCTGTTATCATPAGCATTTTAAATGAAGCCTGG 447
Db 55991 TATGATTATGATAACCTGGCCCAAGACATCCGTTGTCTAGCATTTTAAATGAAGCCTGG 55932
Qy 448 TCAACTCTCTTCCGACAGATTTACAGTGTCTTGTAGACATCCCGGATATCTCTGCTAGAA 507
Db 55931 TCCACACTCTCTTCCGACAGATTTACAGTGTCTTGTAGACATCCCGGATATCTCTGCTAGAA 55972
Qy 508 GAAGTGATCTCTTGTAGATGACTACAGTGTATAGAGACACCTGAGAGGCGCTTGGCCCAAT 567
Db 55871 GAGGTCACTCTTGTAGATGACTACAGTGTATAGAGACACCTGAGAGGCGCTTGGCCCAAG 55812
Qy 568 GAGCTTTCGGGACTGCCCAAGGTGCGCTGATCGCGCCAAACAAGAGAGAGGGGCTTGGTG 627
Db 55811 AGCTGTACACTGCCCTTAGTGTACCTGATCCTGTAGCA--AGAGAGGGGCTTAGTG 55754
Qy 628 CGAGCCCGGCTGTGGGGCGCTGTGCGGAGGGGCGAGTTCGACCTTCTCTGGACTGT 687
Db 55753 CAAG-CCGGCTACTGGGAGCCTCTGGGGCCAGGGGCGAAGTGTGAGGTTTCTGGACTGT 55695
Qy 688 CACTGTGAGTCCACAGAGGGTGGCTGAGCGCGCTGTGACAGAGATCCATGAGAGGAG 747
Db 55694 CACTGTGAGTGTGATGAGGGTGGCTGGAG--TCTGTGACAGAGATCCACGAGAGGAG 55637
Qy 748 TCGGCACTGTGTGCTCCCGGTGATGTGATCGACTGGAACACCTTCGAATACCTGGGG 807
Db 55636 TCGGCGGTGTGTGCTCCCGGTATCGATGTATTGACTGGACACTTCTGTAGTACTTGGGC 55577
Qy 808 AACTCCGGGGAGCCCCAGATCGGCGGTTTCACTGGAGGCTGGTTCACCGTGCAACACA 867
Db 55576 AACTCC-GGAAGCCCCAGATTGGAGGCTTGTGACTGGGCGCTAGTATTACCGTGCAACAGT 55518
Qy 868 GTTCTGAGAGGAGAGATACGATGCAATCCCGCTGATGTATCAGAGTCTCCCAACA 927
Db 55517 GTCCCTTGGCGGAGCGGAAGTGTATGCGGACTCTATTGACATTTACAGGTCTCCAACT 55458
Qy 928 ATGGCTGTGGGCTGTTTGTGTGATGAAGAAATATTTGAATATCTGGGGCTTTATGAT 987
Db 55457 ATGGCTGGGAGCTGTTTGTGTGATGAAGAAATATTTGAATATCTGGGGCTTTATGAT 55398
Qy 988 ACAGAAATGGAAGTTTGGGGAGGAGAAACCTCGAAATTTCTTTAGGATCTGGCAGTGT 1047
Db 55397 ACCGGAATGGAAGTCTAGGGAGGAGAAACCTTGAATTTCTCTTTAGGATCTGGCAATGT 55338
Qy 1048 GGTGGGTTCTGGAAACACACCCATGTTCCCATGTTGGCCATGTTTCCCAAGCAAGCT 1107
Db 55337 GGTGACTCTCTGGAAACACACCCCTTACTCCACGTTGGGCCACGCTTCCCTTAAGCAAGCT 55278
Qy 1108 CCTACTCCCGCAACAAGGCTCTGGCCAAACAGTGTTCGTGAGCTGAAATGATGATGAT 1167
Db 55277 CCTATTACGAGCAAGGCTCTAGCCAAACAGTGTCCGAGTGCAGAGTGTGGCGGAT 55218
Qy 1168 GAAATTAAGAGCTCTACTACATCGCAACCCCGTCCCGCTTGGAACTTTTGGGGAT 1227
Db 55217 GAACTTAAGAAACTCTACACCCCGAAATCCCC-AGCCCGCTTGGAACTTCTGGGGAG 55159
Qy 1228 GTGACAGAGAGGAGGAGCTCGGGACAGCTCCAGTGTAAAGATTTCAAGTGTCTTGT 1287
Db 55158 GTGACAGAGAGGAGGAGGCTTTCAGCCCAAGCTCCAGTGTGAAGATTTAAAGCGGTTCCTG 55099
Qy 1288 GAGACTGTGTATCCAGAACTGCTGCTGCTGAGACAGGCTTCTTCTGGGATGCTC 1347
Db 55098 GCTACACTGTATCCAGAACTGCTGCTGCTGAGGACAGGCTTCTTCTGGGAGGCTT 55039
Qy 1348 CAGAAACAGGACTTAAACAGACTAGTCTGTTTGAATATACCTTCCCGATGAAAAACAGATT 1407
Db 55038 GAGAAACAGGACTTAAACAGACTTAAACAGACTTAAACAGACTTAAACAGACTTAAACAGATT 54979
Qy 1408 GTGGGACACAGGCTTCTGATCTGCTGATGGGATGGGCGGAGTCACTTTTTCGAG 1467
Db 54978 GAGGGCCACAGGCTTCTGATCTGCTGATGGGATGGGCGGAGTCACTTTTTCGAG 54919
Qy 1468 TACAGCTCCAGAGAAATACGCTTAAACACCCACAGCTGAGGGCTGCAATTCGCTGTG 1527
```

Search completed: December 7, 2003, 18:37:52  
Job time : 6377.45 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 7, 2003, 13:31:08 ; Search time 99.1527 Seconds  
(without alignments)  
7772.402 Million cell updates/sec

Title: US-10-074-527-3

Perfect score: 1746

Sequence: 1 atgtggggcgacggcg.....tcaaagagcgcatgttatga 1746

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.\*

- 1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/2/ina/PTUS\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	496.2	28.4	1737	4	US-09-217-306B-1
2	285.6	16.4	1920	3	US-09-376-856-1
3	284	16.3	3332	3	US-09-347-488-1
4	253.8	14.5	1521	4	US-09-795-926-30
5	253.8	14.5	1812	4	US-09-795-926-42
6	253.8	14.5	3896	4	US-09-795-926-44
7	232.2	13.3	1617	2	US-08-967-508-18
8	232.2	13.3	1617	3	US-08-967-506-18
9	232.2	13.3	1617	5	PCT-US94-02552-18
10	232.2	13.3	1680	2	US-08-967-508-11
11	232.2	13.3	1680	3	US-08-967-506-11
12	232.2	13.3	1680	5	PCT-US94-02552-11
13	232.2	13.3	2294	2	US-08-967-508-10
14	232.2	13.3	2294	3	US-08-967-506-10
15	232.2	13.3	2294	5	PCT-US94-02552-10
16	222	12.7	3889	2	US-08-648-298-1
17	201.4	11.5	810	4	US-09-795-926-26
18	201.4	11.5	1101	4	US-09-795-926-38
19	198.8	11.4	1608	4	US-08-795-926-28
20	198.8	11.4	1899	4	US-08-795-926-40
21	134.8	7.7	675	4	US-09-795-926-22
22	134.8	7.7	966	4	US-09-795-926-34
23	97.6	5.6	903	4	US-09-620-312D-938
24	87.4	5.0	322	1	US-08-594-031-61
25	76.4	4.4	262	1	US-08-594-031-62
26	74.2	4.2	2416	4	US-08-620-312D-112
27	69.2	4.0	339	4	US-09-795-926-24

c

28	69.2	4.0	432	4	US-09-795-926-20	Sequence 20, Appl
29	69.2	4.0	630	4	US-09-795-926-36	Sequence 36, Appl
30	69.2	4.0	723	4	US-09-795-926-32	Sequence 32, Appl
31	64.6	3.7	331	3	US-09-328-111-94	Sequence 94, Appl
c 32	63.2	3.6	152331	3	US-09-128-155-16	Sequence 16, Appl
33	63	3.6	319	3	US-09-165-264-8	Sequence 8, Appl
34	60.6	3.5	320	3	US-09-165-264-11	Sequence 11, Appl
35	60.4	3.5	4488	3	US-08-406-030A-3	Sequence 3, Appl
36	59.8	3.4	320	3	US-09-165-264-14	Sequence 14, Appl
c 37	59.4	3.4	12001	1	US-08-458-568A-11	Sequence 11, Appl
38	59.2	3.4	2793	1	US-08-209-747-1	Sequence 1, Appl
39	59.2	3.4	2793	1	US-08-458-298-1	Sequence 1, Appl
40	59.2	3.4	4403765	3	US-09-103-840A-2	Sequence 2, Appl
41	59.2	3.4	4411529	3	US-09-103-840A-1	Sequence 1, Appl
42	58.8	3.4	2214	3	US-08-864-038A-1	Sequence 1, Appl
43	58.8	3.4	3331	3	US-08-864-038A-2	Sequence 2, Appl
44	58.8	3.4	3331	3	US-08-864-038A-4	Sequence 4, Appl
45	58.6	3.4	320	3	US-09-165-264-7	Sequence 7, Appl

#### ALIGNMENTS

##### RESULT 1

US-09-217-306B-1  
; Sequence 1, Application US/09217306B  
; Patent No. 6465220  
; GENERAL INFORMATION:  
; APPLICANT: Hassan, Helle  
; APPLICANT: Clausen, Henrik  
; APPLICANT: Bennett, Eric P.  
; TITLE OF INVENTION: Glycosylation Using Galnac-T4 Transferase  
; FILE REFERENCE: 8850\*1  
; CURRENT APPLICATION NUMBER: US/09/217,306B  
; CURRENT FILING DATE: 1998-12-21  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 1  
; LENGTH: 1737  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-217-306B-1

Query Match		28.4%	Score 496.2;	DB 4;	Length 1737;
Best Local Similarity		62.3%	Pred. No. 2.4e-104;		
Matches		796;	Conservative	0;	Mismatches 478;
				Indels	3;
				Gaps	1;
Qy	220	GCCTGGCGCGCGGGCGAGGCGGTGCGGCTGCAGCTGAGGCGGAGGAGTCCGGCTG	279		
Db	217	GCACCTTGGGAGTGGGGAAAGCCAGCAACTCCAGCTCAACGAGGATGAACCTGAAGCAG	276		
Qy	280	CAGGAGGAGCGTGGCGCTGCACCATTAACATCTACCTCAGCGACCCATCTCACTG	339		
Db	277	CAAGAAGAACTAATGAGAGATAGCCATCAATATTTTACCTCAGTGACGATTTCCCTG	336		
Qy	340	CACCGCGCTGCCGAGCGCTGGAACCCGCTGCAAGAGAGAAATATGATTATGAT	399		
Db	337	CATCGACATAGAGGATAAAGATGATAGTGAAGTCCAGAGATTCACATATAGG	396		
Qy	400	AATTTGCCAGGACATCTGTTATCATAGCATTTTATAAAGCCTGGTCAACTCTCCTT	459		
Db	397	ACACTTCTACCACTCTGTTATCATCTTCTATCAACGAAGCCTGGTGCAGCTTCTGTC	456		
Qy	460	CGGACAGTTTACAGTCTCTTGGAGACATCCCGGATATCTGTAGAGAGATGATCTTT	519		
Db	457	CGTACCATTACAGTGTTTTAGAAACTTCTCTCGAGTCTTTTGAAGAGATCATCTTG	516		
Qy	520	GTAGATGACTACAGTATAGAGAGACCTGAAGAGCGCTTGGCCATAGAGTTCGGGA	579		
Db	517	GTGATGACTTGTAGTGACAGGTTTATTGGAACACACAACTTGAACCTTACATCAGCAAT	576		
Qy	580	CTGCCAAGGTGCGCTGATCCCGGCAACAAAGAGAGCGCTGTGCGAGCCGCGCTG	639		



577 CTGATAGATACGCTTGTATTAGGACCAATAAGCGAGAGGGGCTGTGTAGGGCCGCTG 636  
Db  
QY CTGGGGGCTCTCGGGGAGGGGCGATGCTCTGACCTTCTGGAGCTGCTCACTGTGAGTGC 699  
Db  
637 ATTGGGGCCACTTTCGCCACTGGGAGCGCTCTCACTTCTGTATTTGCTGCTGAGTGT 696  
QY 700 CACGAAGGGTGGCTGGAGCCGCTGTCGAGAGGATCCATGAAGAGGAGTGGCAGTGGTG 759  
Db 697 AATCCGGTGGCTGGACCCGCTTTGGAGAGGATGGAGATATGAACACAGCAGTGTG 756  
QY 760 TGCCCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 819  
Db 757 TGTCTGTATAGACACAAATGATGATGATGATGATGATGATGATGATGATGATGATG 816  
QY 820 CCCAGATCGGGGTTTTCAGATGAGGGGCTGTGTTCACTGCGACACAGTTCCTGAGAG 879  
Db 817 CCCATGATGGTGGGTTTTCAGTGGGTTTAACTTTCAGTGGCATTCCTGCCCCAAG 876  
QY 880 GAGAGGATACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 939  
Db 877 GAAGGGACAGGGGATCAAGAAATGACCCCATCAGATCACTACCATGGCTGGAGGA 936  
QY 940 CTGTTTGTCTGAGTAAGAAATATTTGAAATATCTGGGGTCTTATGATACAGGAATGAA 999  
Db 937 CTGTTTGTCTGAGTAAGAAATATTTGAAATATCTGGGGTCTTATGATACAGGAATGAA 996  
QY 1000 GTTTGGGAGGAGAAACCTCGAATTTTCTTTAGGATCTGCGAGTGTGGGGTCTG 1059  
Db 997 GTTGGGAGGAGAAACCTGAGCTGCTTTTGGGGTGTGGGAGTGTGGGCAATG 1056  
QY 1060 GAAACACACCCATGTTTCCCATGTTGGCCATGTTTCCCAAGCAAGCTTCCCTACTCCCGC 1119  
Db 1057 GAGATCCACCGGTGTTCCCATGTTGGCCATGTTTCCCAAGCGGGACCATATGCTCGC 1116  
QY 1120 AACAGGCTCTGGCCACACAGTGTGTCGACGCTGAAGTATGATGATGATGATGATGATG 1179  
Db 1117 CCCAATTTCTACAGAAATCTGCTCGGGCAGCAGAAATTTTGGATGATGATGATGATG 1176  
QY 1180 CTCTACTACCATCGAACCCCGTGGCCGCTTGGACCTTTTGGGATGTCACAGAGAG 1239  
Db 1177 CACTTCTACATAGAAACCTCCAGCAAGAAAGAGCTTATGTTGATGATGATGATGATG 1236  
QY 1240 AAGCAGCTCCGGGACAAAGCTCCAGTGTAAAGACTTCAAGTGTGTTCTTGGAGCTGTG 1299  
Db 1237 AATATTACTACGAGCGGTGAGATGCAAGACTTGTGATGTTTGAATAACGTTTTT 1296  
QY 1300 CCAGAACTGATGTCCTGAGACAGGCTGCTGCTTCTTGGGATGCTCCAGAACAAAGGA 1359  
Db 1297 CCTAATTTTACATGTTTCCAGAGGATAGACCGCTGGCATGGGGCTATTTCGCAGTAGAG 1356  
QY 1360 CTACAGACTACTGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1419  
Db 1357 ATCTCGTGTGAATGTTTGAATGATGATGATGATGATGATGATGATGATGATGATG 1414  
QY 1420 GTCAATCTGTACTCTGTGATGGGATGGCGAGAAATCAGTGTGTTTTCGAGTACACGTC 1479  
Db 1415 -TTTCACTGTTTGGATGATGATGATGATGATGATGATGATGATGATGATGATG 1473  
QY 1480 AAGAAATACGCTATAA 1496  
Db 1474 AAGAAATAAGGTTTAA 1490

RESULT 2

US-09-376-856-1  
; Sequence 1, Application US/09376856  
; Patent No. 6235510  
; GENERAL INFORMATION:  
; APPLICANT: KUMAR, SANJAY  
; APPLICANT: VAN HORN, MARION M.  
; APPLICANT: LARK, MICHAEL  
; TITLE OF INVENTION: ppGantase-T6  
; FILE REFERENCE: GP-70638

; CURRENT APPLICATION NUMBER: US/09/376,856  
; CURRENT FILING DATE: 1999-08-18  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 1920  
; TYPE: DNA  
; ORGANISM: HOMO SAPIENS  
US-09-376-856-1

Query Match 16.4%; Score 285.6; DB 3; Length 1920;  
Best Local Similarity 53.9%; Pred. No. 3.5e-36;  
Matches 741; Conservative 0; Mismatches 599; Indels 36; Gaps 6;

QY 319 CTGAGGACCGCATCTCACTGCACCCGCGCTCCGAGCGCTGGAAACCGCTGTGCAAA 378  
Db 487 CTGAGTGGCCGCAATCCCTTCCAGAGGGCTCTCCGAGGTGGGACACCCACTGTG---T 543  
QY 379 GAGAAGAAATATGATTATGATAATTTGCCAGGACATCTGTATCATAGCAATTTTAAAT 438  
Db 544 CTGACGACGACCCCTCAGGACAGCCTGCCACAGCCAGCGTCATCTCTGTTTCCATGAT 603  
QY 439 GAAGCCTGTCACACTCTCTTCGACAGATTACAGTGTCTTTCAGACATCCCGGATATC 498  
Db 604 GAGGCTGTGTCCTCTCTGCGGACTGTACACAGCATCTCTGACAGTGTGCCAGGGCC 663  
QY 499 CTGTAGAAGAGTGTCTCTGTAGATGACTACAGTGTATAGAGACCTTGAAGAGCGC 558  
Db 664 TTCTGAAGAGATCATCTCTGAGGACCTCAGCAGCAAGACAACTCAAGTCTGCT 723  
QY 559 TTGGCCAAATGAGCTTTTGGGACTGCCCAAGGTGCGCTGTATCCGCGCAACAAGAGAG 618  
Db 724 CTGACGCAATATGTCGACAGCTGGAGGGGTGAAGTTACTCAGGAGCAACAAGAGGCTG 783  
QY 619 GGCTGTGTGAGCGCCGCTGTCTGGGGGCTCTGCGCGAGGGGCGATGTTCTGACCTTC 678  
Db 784 GGTGCCATCAGGCGCCGAGTGTCTGGGGGCCACAGAGCCACCGGGGATGTCTGCTTTC 843  
QY 679 CTGAGCTGTCACTGTAGTCCGACAGAGGTGCTGGAGCGCTGTCTGACAGAGATCCAT 738  
Db 844 ATGATGCCCCACTGCGAGTGCACCCAGGCTGCTGGAGGCCCTCTCTCAGCAGATAGCT 903  
QY 739 GAAGAGAGTCCGAGTGTGTCGCGGTGATGATGATGATGATGATGATGATGATGATG 798  
Db 904 GGTGACAGGAGCGGAGTGTATCTCCGGTGTATGATGATGATGATGATGATGATGATG 963  
QY 799 TACCTGGGAACTCCGCGGAGCCAGATCGCGGTTTCAGCTGGAGGCTGTGTTTCAAG 858  
Db 964 TA---TTTACCCCTCAAAGGACCTGCAGCGTGGGGTGTGAGCTGGAAGTGTTCAC 1020  
QY 859 TGGCACACAGTCTCTGAGAGGAGAGATACGGATGCAATCCCGCTCGATGTCTATCAGG 918  
Db 1021 TGGGAACCTTTCAGAGCATGTGAGGAAGCCCTCAGTCCCGCCATAGCCCCATCAGG 1080  
QY 919 TCTCCAAATATGGCTGTGGGCTGTTTGTCTGTGATGAAGAAATATTTTGAATATCTCG 978  
Db 1081 AGCCTGTGTGTCGCGAGAGGTGTGGCCATGACAGACATTAATTTCCAAAACACTGGA 1140  
QY 979 TCTTATGATACAGGAATGGAAGTGTGGGAGGAGAAACCTCGAATTTCTCTTTAGATC 1038  
Db 1141 GCGTATGACTCTTATGTCGCTGCGAGGTGGTGAAGAACCTCGAATCTGTCTTCAAGGCC 1200  
QY 1039 TGGCAGTGTGGGGTCTTGGAAACACACCCATGTTTCCCATGTTGGCCATGTTTTCGCC 1098  
Db 1201 TGGCTGTGTGGTCTGTGTAATCTTCCCTGCTCTCGGTGAGACATCTACCA 1260  
QY 1099 AAGCAAGCTCCCTACTCCCGCAACAGGCTCTGSCC-----AACAGTGTTCGTGCA 1149  
Db 1261 AATCAGGATTTCCCAATTCCTCCCTGACAGAGGCCACCTCGAGGAACAGGGTTCGCA 1320  
QY 1150 GCTGAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1209  
Db 1321 GCTGAGACCTGGCTGGGGTCAATTCAAAAGAAACCTTCTTACAAGCATGCCCCAGAGGCTTC 1380





```
Db 2096 GAGGAAGCGCTGGCCATCCACAGCAGACTGGGACTTCAGGAGAAATGGGATGATGTC 2155
Qy 1621 CACGAACAGTCCAGAATAATGTCTCAGCGTGGAGGAGAGTGCAGTGCAGATTT 1676
Db 2156 CACATTTCTTGGGAAATGCAATGCAAGACTGTGTGTCAGAGAAAAACAATAAGATTT 2211

RESULT 4
US-09-795-926-30
; Sequence 30, Application US/09795926
; Patent No. 6555669
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Hilbun, Erin
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; APPLICANT: Walke, D. Wade
; APPLICANT: Wilganowski, Nathaniel L.
; APPLICANT: Hu, Yi
; APPLICANT: Kieke, James Alvin
; APPLICANT: Potter, David George
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
; FILE REFERENCE: LEX-0144-USA
; CURRENT APPLICATION NUMBER: US/09/795,926
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/185,920
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 60/186,558
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/191,849
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 1521
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-795-926-30

Query Match 14.5%; Score 253.8; DB 4; Length 1521;
Best Local Similarity 55.5%; Pred. No. 5.9e-49;
Matches 561; Conservative 0; Mismatches 432; Indels 18; Gaps 3;

Qy 308 TTAACATCTAGCTCAGCGACCGCATCTCACTGCGACCGCGCTGCGGAGCGCTGGAAACC 367
Db 47 TTAACATCTAGCTCAGTGATAAATCTCTTGAATGCTCTCTCCAGATATCGGCACC 106
Qy 368 CGCTGTGGAAGAGAAATATGATTAATGATAATTTGGCCAGGACATCTGTTATCATAG 427
Db 107 CAAATCTGCAACAGCAAGCGCTACCT--GGAGACACTTCCCAACACAGCATCATCTCC 163
Qy 428 CATTTTAAATGAAGCGTGTCAACTCTCCCTCGGACAGTTTACAGTGTCTCTGAGACAT 487
Db 164 CTTTCCAAACAGAGGGCTGGTCTCTCCCTCCCGACCGTCCACAGTGTGCTCAATCGCT 223
Qy 488 CCCCAGATATCTGCTAGAAGATGATCTCTTGTAGATGACTACAGTGTATAGAGAGCACC 547
Db 224 CGCTCCAGAGCTGCTGCCGAGATGTTGTTGTCGACGACTTCACTGATCGAGAGCACC 283
Qy 548 TGAAGAGCGCTTGGCCAAATGAGCTTTCGGGACTGCCCCAAGTGCCTGATTCGCGCCA 607
Db 284 TGAAGAGCGCTTGAAGACTACATGCGCCCTTTTCCCGAGTGTGAGGATTTCTTGAACCA 343
Qy 608 ACAAGAGAGCGCTGTGTGAGCGCGCTGCTGGGGCGCTCTCGCGAGGGCGCATG 667
Db 344 AGAACCGGAAGGGCTGATAGAGACCCGAAATGCTGGGGCGCTCACTGGCAACTGGGATG 403
Qy 668 TTCTGACCTTCTGACTGTCACTGTGAGTGCACGAAGGGTGGCTGGAGCGCTGTGCTGC 727
```

```
Db 404 TCATCACAATTTCTGGATTCACACTGTGAAGCCAATGTCAACTGGCTTCCCCCTTGGTTG 463
Qy 728 AGAGGATCCATGAAGAGGAGTCCGAGTGGTGGCCGGTGGATTTGATGTGATCGACTGGA 787
Db 464 ACCGATTTGCTCGGAACCGCAAGCATTGTGTGCCCCGATGATGATGATTAATGACCATG 523
Qy 788 ACACCTTTGAAATACCTCGGGGAACTCCGGGGAGCCCGAGATCGCGGGTTTCGACTGGAGGC 847
Db 524 ACACCTTTGCTGACGAGACACAGCAGGGGATGCCATGCGGGGAGCCTTTGACTGGGAGA 583
Qy 848 TGGTGTTCAGTGGCACAAGTTTCTCGAGAGGAGAGGATACCGATGCAATCCCGCTGCG 907
Db 584 TGTACTACAAGCGGATCCCGATCCCT-----CCAGAACTGCAGAAAGCTGACCCAGCG 637
Qy 908 ATGTCAATCAGTCTCCCAACAAATGGCTGGTGGGCTGTTTCTGTGAGTAAGAAATATTTT 967
Db 638 ACCATTTGAGTCTCCCGTGTATGCGCGGTGGAATGTTTCCCGTGGATCGGAAGTGGTTCT 697
Qy 968 AATATCTGGGGTCTTATGATACAGGAATGGAAGTTTGGGGGAGGAGAAACCTCGAATTTT 1027
Db 698 GGGAACTCGCGGGTATGACCCAGGCTTGAGATCTGGGAGGGGAGCAGTATGAATCT 757
Qy 1028 CTTTAGGATCTGGCAGTGTGGGTTCTGGAACACACACCATGTTCCCATGTTGGCC 1087
Db 758 CTTTCAAGGTGTGATGTGTGGGGCGCATGAGGACATCCCTGCTCCAGGGTGGGCC 817
Qy 1088 ATGTTTTCCCAAGCAAGCTCCCTAC-----TCCCGCAACAGGCTCTGGCCAAACA 1138
Db 818 ATATCTACAGGAAGTATGTGCCCTCAAGGTCCCGCGGAGTCAAGCTGGCCCGGAAACC 877
Qy 1139 GTGTTGCTGAGCTGAAGTATGATGATGAAATTTAAAGAGCTCTACTACCATCGCAACC 1198
Db 878 TTAAGCGGTGGCGAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 937
Qy 1199 CCGTGTCCCGCTTGGAACTTTTGGGATGTGACAGAGAGGAGAGCAGCTCGGGACAAGC 1258
Db 938 CTGAATACCGCCACCTCTCCGTGGGATGTGCGAGTCCAGAAAAAGCTCCGAGCTCCC 997
Qy 1259 TCCAGTGTAAAGACTTCAGTGGTCTTGGAGACTGTGATCCAGACTGC 1309
Db 998 TTAACCTGCAAGAGTTCAGTGGTTCATGAGAGATAGCTGGGACCTGC 1048

RESULT 5
US-09-795-926-42
; Sequence 42, Application US/09795926
; Patent No. 6555669
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Hilbun, Erin
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; APPLICANT: Walke, D. Wade
; APPLICANT: Wilganowski, Nathaniel L.
; APPLICANT: Hu, Yi
; APPLICANT: Kieke, James Alvin
; APPLICANT: Potter, David George
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
; FILE REFERENCE: LEX-0144-USA
; CURRENT APPLICATION NUMBER: US/09/795,926
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/185,920
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 60/186,558
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/191,849
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 47
```

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 42

; LENGTH: 1812

; TYPE: DNA

; ORGANISM: homo sapiens

US-09-795-926-42

Query Match 14.5%; Score 253.8; DB 4; Length 1812;

Best Local Similarity 55.5%; Pred. No. 6.3e-49;

Matches 561; Conservative 0; Mismatches 432; Indels 18; Gaps 3;

```
Qy 308 TTAACATCTACTCAGCGACCGCATCTCACTGCACCGCGCTCCCGAGCGCTGGAAACC 367
Db 338 TTAACATCTAGTCAGTGATTAATCTCTTGAATCGCTCTCTCCAGATATCCGGCAACC 397
Qy 368 CGCTGTGCAAGAGAGAAATATGATTATGATAATTTGCCAGGACATCTGTTATCATAG 427
Db 398 CAAACTGCAACAGCAAGCGCTACCT--GGAGACACTTCCCAACACAGCATCATATCC 454
Qy 428 CATTTTATATGAAGCCTGTCAACTCTCTCTCGGACAGTTTACAGTGTCTTGAGACAT 487
Db 455 CTTTCCAACAGAGGCTGTGCTCTCTCTCCGACCGTCCACAGTGTGCTCAATCGCT 514
Qy 488 CCCGGATATCTCTAGAGAGATGATCTCTGTAGATGACTACAGTATAGAGACACC 547
Db 515 CGCTCCAGAGCTGTGCGCGAGATTGACTGTGTCGACACTTCAGTGATCGAGAGACCC 574
Qy 548 TGAAGAGCGCTTGGCCAAATGAGCTTTCGGGACTGCCCAAGGTGCGCTGATCCGCGCCA 607
Db 575 TGAAGAGCTCTTGAAGACTACATGCGCTTTTCCCGAGTGTGAGGATCTTCGAACCA 634
Qy 608 ACAAGAGAGAGGCTCTGTGCGACCGCGCTGTCTGGGGCGTCTGCGCGAGGGCGCATG 667
Db 635 AGAAACGGGAAGGCTGATAAGGACCCGGAATGCTGGGGGCTCAGTGGCAACTGGGATG 694
Qy 668 TTCTGACCTTCTCGACTGTACACTGTAGTCCACGAGGCTGCTGAGCGGTGCTGTC 727
Db 695 TCATCAATCTTGGATTACACTGTGAAGCAATGTCAACTGGCTTCCCGCTTGTGTTG 754
Qy 728 AGAGGATCCATGAAGAGAGTGGCGAGTGTGTGCGGCTGATTGATGTGACTGGA 787
Db 755 ACCGATTTGCTCGGACCGCAAGACCATGTGTGCGCGATGATTGATGTAATTGACCATG 814
Qy 788 ACACCTTTCGAATACCTGGGAACTCCGCGGAGCCCCAGATCGCGGTTTGACTGGAGGC 847
Db 815 ACGATTTTCGATACGACACACAGGACAGGGATGCCATCGCGGGAGCCTTTGACTGGAGA 874
Qy 848 TGGTGTTCAGTGGCACACAGTTCTTGAGGGAGAGATACGATGCAATCCCGCTGCG 907
Db 875 TGTACTACAAGCGGATCCCGATCCCT-----CCAGAACTGCAGAAAGCTGACCCAGCG 928
Qy 908 ATGTTCATCAGTCTCCAAACAATGCTGTGGGCTGTTTGTCTGCTAGTAAGAAATATTTTG 967
Db 929 ACCATTTTGATCTCCCGTATGTCGCGGTGACTGTTCGCGTGGATCGGAATGGTTCT 988
Qy 968 AATATCTGGGCTTTATGATACAGGAATGGAATTTGGGAGGAGAAACCTCGAATTTT 1027
Db 989 GGGAACTCGGCGGATATGACCCAGGCTTGGAGATCTGGGAGGGGAGCAGTATGAAATCT 1048
Qy 1028 CCTTTAGGATCTGCACTGTGGTGGGTTCTGGAAACACACCCATGTTCCATGTTGGCC 1087
Db 1049 CCTTCAAGGTGTGATGTGTGGGGCGCCATGAGGAGCATCTCCCTGTCTCAGGGTGGGCC 1108
Qy 1088 ATGTTTTTCCCAAGCAAGCTCCCTAC-----TCCCGCAACAAGCTCTGCGCAACA 1138
Db 1109 ATATCTACAGGAATGATGTGCCCTTACAGGTCCCGCGGAGTCAAGCTGCGCGCAACC 1168
Qy 1139 GTGTTCGTGCACTGAAAGTATGGATGGATGAATTTAAAGAGCTCTACTACCATCGCAACC 1198
Db 1169 TTAAGCGGGTGGCGAAGTGTGGATGATGATGATGATGATGATGATGATGATGATGATG 1228
Qy 1199 CCGTGCCTGCTTGGAACTTTTGGGATGTGACAGAGAGGAAGCAGCTCCGGGACAAGC 1258
```

```
Db 1229 CTGAATACCGCCACCTCTCCGCTGGGGATTCGCGAGTCCAGAAAAAGCTCGGAGCTCCC 1288
Qy 1259 TCCAGTGTAAAGACTTTCAGTGTGTTCTTTGGAGACTGTGTATCCAGAACTGC 1309
Db 1289 TTAACGCAAGAGTTTTCAGTGTGTTTATGACGAAGATAGCTGGGACCTGC 1339
```

#### RESULT 6

US-09-795-926-44

; Sequence 44, Application US/09795926

; Patent No. 6555669

; GENERAL INFORMATION:

; APPLICANT: Donoho, Gregory

; APPLICANT: Hilbun, Erin

; APPLICANT: Turner, C. Alexander Jr.

; APPLICANT: Friedrich, Glenn

; APPLICANT: Abuin, Alejandro

; APPLICANT: Zambrowicz, Brian

; APPLICANT: Sands, Arthur T.

; APPLICANT: Walke, D. Wade

; APPLICANT: Wilganowski, Nathaniel L.

; APPLICANT: Hu, Yi

; APPLICANT: Kieke, James Alvin

; APPLICANT: Potter, David George

; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND

; FILE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME

; FILE REFERENCE: LEX-0144-USA

; CURRENT APPLICATION NUMBER: US/09/795,926

; PRIOR FILING DATE: 2001-02-28

; PRIOR APPLICATION NUMBER: US 60/185,920

; PRIOR FILING DATE: 2000-02-29

; PRIOR APPLICATION NUMBER: US 60/186,558

; PRIOR FILING DATE: 2000-03-02

; PRIOR APPLICATION NUMBER: US 60/191,849

; PRIOR FILING DATE: 2000-03-24

; NUMBER OF SEQ ID NOS: 47

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 44

; LENGTH: 3896

; TYPE: DNA

; ORGANISM: homo sapiens

US-09-795-926-44

Query Match

Best Local Similarity 55.5%; Pred. No. 8e-49;

Matches 561; Conservative 0; Mismatches 432; Indels 18; Gaps 3;

```
Qy 308 TTAACATCTACTCAGGACCGCATCTCACTGCACCGCGCTGCGCGAGCGCTGGAAACC 367
Db 347 TTAACATCTACGTACGTGATGATAAATCTCTTGAATCGCTCTCTCCAGATATCCGGCACC 406
Qy 368 CGCTGTGCAAGAGAGAAATATGATTATGATAATTTGCCAGGACATCTGTTATCATAG 427
Db 407 CAACTGCAACAGCAAGCGCTACCT--GGAGACACTTCCCAACACAGCATCATATCC 463
Qy 428 CATTTTATATGAAGCCTGTGCAACTCTCTCTCGGACAGTTTACAGTGTCTTGAGACAT 487
Db 464 CTTTCCACACAGAGGCTGTCTCTCTCTCTCCGACCGTCCACAGTGTCTCAATCGCT 523
Qy 488 CCCGGATATCTCTGTAGAAAGAGTGTATCTCTGTAGATGACTACAGTGTAGAGAGCACC 547
Db 524 CGCTCTCAGAGCTGTGTCGCGAGATGTTACTGTGTCGACGACTTCAGTGTACGAGAGCACC 583
Qy 548 TGAAGAGCGCTTGGCCAAATGAGCTTTTCGGAGCTGCGCAGGTGGCGCTGATCCGCGCCA 607
Db 584 TGAAGAGCGCTTGAAGACTCATGGCCCTTTTCCCGAGTGTGAGGATTTCTTGAACCA 643
Qy 608 ACAAGAGAGAGGCGCTGGTCCGAGCGCGCTGTCTGGGGGCGTCTGCGCGAGGGCGCATG 667
Db 644 AGAAACGGGAAGGCTGATGAAGACCCGATGCTGGGGGCTCAGTGGCAACTGGGATG 703
Qy 668 TTCTGACCTTCTGGAATCTTTGGGATGTGACAGAGAGGAAGCAGCTCCGGGACAAGC 727
```

Db 704 TCATCACATCTCTGGATTACACTGTGAAGCAATGTCACTGGCTTCCCTTGGCTTG 763  
Qy 728 AGAGATCCATGAAGAGAGTGGCAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 787  
Db 764 ACCGATTGCTCGGAACCGCAAGACCATGTGTGCGCGATGATTGATTGATTGATTG 823  
Qy 788 ACACCTTCGATATCTCGGGAACCTCGGGAAGCCAGATCGGCGGTTTCGACTGGAGGC 847  
Db 824 ACGACTTCGTACGACACAGCAGGAGGATGCCATCGGGGAGCCCTTTCGACTGGGAGA 883  
Qy 848 TGGTGTTCACGTGGCAGACAGTTCCTGAGAGGAGAGATACGATGCAATCCCGCTGCG 907  
Db 884 TGTACTACAGCGGATCCGATCCCT-----CCAGAACTCGAGAAAGCTGACCCAGCG 937  
Qy 908 ATGTCATCAGCTCCCAACATGCTGTGGGCTGTGCTGTGCTGAGTAAGAAATTTTG 967  
Db 938 ACCATTGTGATCTCCCGTGTGCGCGGTGACTGTTCGCGGTGATCGGAAGTGGTCT 997  
Qy 968 AATATCTGGGCTTATGATACAGGAATGGAAGTTTGGGAGGAGAGAAACCTCGAATTTT 1027  
Db 998 GGAACCTCGCGGATGACCCAGGCTTGGAGATCGGGAGGGGAGCAGTATCAATCT 1057  
Qy 1028 CTTTAGATCTGGCAGTGGTGGTCTGGAACACACACCATGTTCCATGTTGGCC 1087  
Db 1058 CTTCAAGGTGTGGTGTGGGCGCGCATGAGGAGCATCCCTGTCTCCAGGCTGGGCC 1117  
Qy 1088 ATGTTTCCCAAGCAGCTCCCTAC-----TCCGCAACAAGGCTCTGCCCAACA 1138  
Db 1118 ATATCTACAGGAATGATGTCCTTACAGGTCCCGCGCGGAGTCAGCGTGGCCCGGAACC 1177  
Qy 1139 GTGTCTGTGACGTGAAGTATGGATGGAATTTAAAGAGCTCTACTACCATCGCAACC 1198  
Db 1178 TTAAGCGGTGCGCGAAGTGTGGATGATGATGATGATGATGATGATGATGATGATG 1237  
Qy 1199 CCGTGGCCGTGGAACTTTTGGGATGTGACAGAGAGAGAGAGAGAGAGAGAGAGAG 1258  
Db 1238 CTGAATACCGCACCTCTCCGTGGGATGTGCGAGTCCAGAAAGCTCGCAGCTCCC 1297  
Qy 1259 TCCAGTGAAGACTTCAAGTGGTCTTGGAGAGCTGTATCCAGACTGC 1309  
Db 1298 TTAAGTGAAGAGTTTCAAGTGGTATTGACGAAGATAGCTGGGACCTGC 1348

## RESULT 7

US-08-967-508-18  
; Sequence 18, Application US/08967508  
; Patent No. 5910570  
; GENERAL INFORMATION:  
; APPLICANT: The Upjohn Company  
; APPLICANT: FOR U.S. PURPOSES ONLY: Elhammer, Ake P. and Homa, Fred L.  
; TITLE OF INVENTION: A Cloned DNA Encoding a UDP-GalNAC:  
; TITLE OF INVENTION: Polypeptide, N-Acetylgalactosaminyltransferase  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pharmacia & Upjohn Company, Intellectual  
; ADDRESS: Property Legal Services  
; STREET: 301 Henrietta Street  
; CITY: Kalamazoo  
; STATE: Michigan  
; COUNTRY: USA  
; ZIP: 49001  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/967,508  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/602,830  
; FILING DATE: 13 No. 5910570ember 1995

; ATTORNEY/AGENT INFORMATION:  
; NAME: Darnley Jr., James D.  
; REGISTRATION NUMBER: 33,673  
; REFERENCE/DOCKET NUMBER: 4755. P CP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 616-833-2210  
; TELEFAX: 616-833-8897  
; TELEX: 224401  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1617 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cdna  
; US-08-967-508-18

Query Match 13.3%; Score 232.2; DB 2; Length 1617;  
Best Local Similarity 54.8%; Pred. No. 5.2e-44;  
Matches 612; Conservative 0; Mismatches 463; Indels 42; Gaps 6;

Qy 392 ATTATGATATTTGCCAGGACATCTCTTATCATAGCATTTTATAATGAAGCCTGGTCAA 451  
Db 269 ATCCAGATAACCTTCCTACAAACAGTGTGGTGTGTTTCCCAATGAGGCTTGGAGCA 328  
Qy 452 CTCTCTTCGACAGTTTACAGTGTCTTCCAGACATCCCGGATATCTCTGTAGAAAG 511  
Db 329 CATTCTGCAATGTCCATAGCGTCATTAATCGCTCACCAGGCACATGCTAGAGAAA 388  
Qy 512 TGATCTCTGTAGATGACTACAGTGTAGAGAGACCTGGAAGGAGCGCTTGGCCAAATGAGC 571  
Db 389 TTGTTCTAGTAGATGATGCGCAGTGAAGAGACTTTTAAAGAACCTCTTAGAGAGTTACG 448  
Qy 572 TTTGGGAGCT---GCCAAGTGGCTGTATCGCGCAACAAGAGAGAGAGGCTGGTGC 628  
Db 449 TGAATAAATTTAAAGTACCCGTTTCAAGTCAATTCGAATGGAGAGCGTCTGGATTTGATCA 508  
Qy 629 GAGCGCGCTGCTGGGCGCTCTCGCGGAGGGCGGATGTTCTGACCTCTCTGACCTGTC 688  
Db 509 GAGCTAGTTAAAGGTGCTGCTGTCTTAAAGCCAGTGTATCACCCTTTTATAGACCGC 568  
Qy 689 ACTGTGAGTCCCAAGAGGTGCTGGAGCGCTGTGTCAGAGAGATCCATGAAGAGAGT 748  
Db 569 ACTGTGAGTGCACAGTGGGCTGGAGCGCTCTCTTAGCAGGATCAAAACATGACAGGA 628  
Qy 749 CGGAGTGGTGTGCGCGTGTATGATGTGATGCTGGAACACCTTCCGAATACCTGGGA 808  
Db 629 AGACAGTGGTCTGTCCCATCATAGATGTGATGATGATGATGATGATGATGATGATG 688  
Qy 809 ACTCCGGGAGCGCCAGATCGGCGGTTTCGACTGGAGGCTGGTGTTCACGTTGGCAGCAG 868  
Db 689 GTTCTGACATGACCTA---TGGCGGTTCACTGGAAGCTCACTTTCGCTGGTATCCTG 745  
Qy 869 TTCTGTAGAGGAGAGG---ATACGGATGCAATCCCGCTGATGTGATCATCAGGTCTCAA 925  
Db 746 TTCCCAAGAGAAATGGACAGAAAGGTGATCGGACTTCTCTGTGTGAGAACACCTA 805  
Qy 926 CAATGGCTGTGGCTGTTTGTGTGTGATGAAGAAATATTTTGAATATCTGGGGTCTTATG 985  
Db 806 CAATGGCAGGAGGCTTTTTCATATAGACAGAGATTACTTTCAGGAAATTTGGAAACA 865  
Qy 986 ATACAGGAATGGAAGTTTGGGAGGAGAAACCTCGAATTTTCTTTAGGATCTGGCAGT 1045  
Db 866 ATGCTGGAATGGATATTTTGGGAGGAGAAACCTAGAAATTTCTTTAGGATTTGGCAGT 925  
Qy 1046 GTGGTGGGTTCTGGAAACACACCCATGTTTCCCATGTGGCCATGTTTTCCTCCCAAGCAAG 1105  
Db 926 GTGGAGGAACCTTTGGAGATGTTTACTTGTCTCACATGTTGGACATGTGTCTTCGGAAGCTA 985  
Qy 1106 CTCCCTACTCCGCAACAAGCTCTGCGCAACAGTGT-----TCGTGCGAG 1150  
Db 986 CACCTTACAGTTCTCCAGGAGGACCGGGGAGATTATCAATAAATAAACAGACGACTTG 1045

QY 1151 CTGAAGTATGATGGATCAATTTAAAGAGCTCTACTACCATCGCAACCCCGTCCCGCT 1210  
Db 1046 CAGAAGTATGATGGATGAATTTCAAGAAATTTCTTCTATATAATTTCTCCAGGTGTACAA 1105  
QY 1211 TGAACCTTTTGGGGATGTGACAGAGAGGAGCAGCTCCGGGACAAGCTCCAGGTGTAAG 1270  
Db 1106 AGGTAGATTAGAGATATATCATCAAGACTTGGTCTAAGGCACAACTCCAATGCAGAC 1165  
QY 1271 ACTTCAAGTGGTCTTGGAGACTGTGTATCCAGAACTGTCATGTCCTGAGACAGGCGTG 1330  
Db 1166 CATTCCTCTGGTACTAGAGAAATTTTATCTGATTTCTAGATTCTCGTCACTATTCT 1225  
QY 1331 GCTTCTTCGGGATCTCCAGAACAAAGACTAAACAGACTACTGTTTGACTATACCTC 1390  
Db 1226 CTTTGGGAGAGATAC-----GAAATGTGGAACAAATCAAGTGTCTAGATAACATGCTA 1279  
QY 1391 CCGATGAAACCAAGATTGTGGGACACCAGGTCAATCTGTACTCTGTATGGGATGGCC 1450  
Db 1280 GAAAGAGATGAAAGTTGGA-----ATTTTAACTGTCTATGGATGGAG 1327  
QY 1451 AGAATCAGTTTTCGAGTACACAGTCCAGAAAGAAAT 1487  
Db 1328 GTAATCAGGTTTCTCTTACACTGCCAACAAAGAAAT 1364

## RESULT 8

US-08-967-506-18  
; Sequence 18, Application US/08967506  
; Patent No. 6096512  
; GENERAL INFORMATION:  
; APPLICANT: The Upjohn Company  
; APPLICANT: FOR U.S. PURPOSES ONLY: Elhammer, Ake P. and Homa, Fred L.  
; TITLE OF INVENTION: A Cloned DNA Encoding a UDP-GalNAC:  
; TITLE OF INVENTION: Polypeptide, N-Acetylgalactosaminyltransferase  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pharmacia & Upjohn Company, Intellectual  
; ADDRESSEE: Property Legal Services  
; STREET: 301 Henrietta Street  
; CITY: Kalamazoo  
; STATE: Michigan  
; COUNTRY: USA  
; ZIP: 49001  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/967,506  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/602,830  
; FILING DATE: 13 No. 6096512ember 1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Barnley Jr., James D.  
; REGISTRATION NUMBER: 33,673  
; REFERENCE/DOCKET NUMBER: 4755.P CP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 616-833-2210  
; TELEFAX: 616-833-8897  
; TELEX: 224401  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1617 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cdna  
US-08-967-506-18

Query Match 13.3%; Score 232.2; DB 3; Length 1617;

Best Local Similarity 54.8%; Pred. No. 5.2e-44;  
Matches 612; Conservative 0; Mismatches 463; Indels 42; Gaps 6;  
QY 392 ATTATGATTAATTTGCCAGGACATCTGTTATCATAGCATTTTATAAGAGCCCTGGTCAA 451  
Db 269 ATCCAGATACTTCTTACCAACAGGTGGTGAATGTTTCCCAATAGAGCTTGGAGCA 328  
QY 452 CTCTCTCTCGGACAGTTTACAGTGTCTTGGAGACATCCCGGATATCTCTCTAGAAAG 511  
Db 329 CACTTCTCGAACTGTCCATAGCGTCATTAATCGTCCACCAAGGCACATGCTAGAGAAA 388  
QY 512 TGATCTTGTAGATGACTACAGTATAGAGAGCACCTGAAGGAGCGCTTGGCCATAGC 571  
Db 389 TTGTTCTAGTAGATGATGCCAGTGAAGAGACTTTTAAAGAGACCTCTAGAGAGTTACG 448  
QY 572 TTTCCGGACT---GCCAAGGTGGCTGATCCGCGCAACAAGAGAGAGAGCGCTGGTGC 628  
Db 449 TGAAGAAATTTAAAGTACCCTGTCATTCGATGGAGCAGCGTCTGGATTGATCA 508  
QY 629 GAGCCCGGTGCTGGGGCGTCTGCGCGAGGGCGGATGTTCTGACCTTCTCGACTGTC 688  
Db 509 GAGTAGGTTAAAGAGGTGCTGTGCTAAAGGCCAAGTATCACCTTTTACGCGC 568  
QY 689 ACTGTGAGTCCACGAGGAGGTGCTGAGGCGCTGTCGAGAGGATCCATGAAGAGAGT 748  
Db 569 ACTGTGAGTGCACAGTGGGGTGGCTGAGGCTCTCTTAGCCAGGATCAAAATGACAGGA 628  
QY 749 CGGAGTGGTGTCCCGGTGATTTGATGATGATCGACTTGAACACCTTCCGAATACCTGGGA 808  
Db 629 AGACAGTGTCTGCTCCATCATAGATGATGATGATGATGATGATGATGATGATGATG 688  
QY 809 ACTCCGGGGAGCCCGAGATCGGCGGTTTTCGACTGGAGGCTGGTGTTCACGCTGGCACACAG 868  
Db 689 GTTCTGACATGACCTA---TGGCGGGTTCAACTTGAAGCTCAACTTTCGCTGGTATCTG 745  
QY 869 TTCTGAGAGGAGAGAG---ATACGATGCAATCCCGGTGATGATGATGATGATGATGATG 925  
Db 746 TTCCCAAGAGAGAAATGGACAGAGGAAAGGTGATCGGACTCTTCTGTGAGAACACCTA 805  
QY 926 CAATGCTGTGGTGTGTTGCTGTGATGATGATGATGATGATGATGATGATGATGATGATG 985  
Db 806 CAATGCGAGGAGGCTTTTCAATGACAGAGATTACTTTCAGGAAATTTGGAACATATG 865  
QY 986 ATACAGGAATGGAAGTTTGGGAGGAGAGAAACCTCGAATTTTCTTTAGGATCTGGCAGT 1045  
Db 866 ATGCTGGAATGGATATTTGGGAGGAGAGAAACCTAGAAATTTCTTTAGGATTTGGCAGT 925  
QY 1046 GTGTTGGGTTCTGGAAACACACCCATGTTCCCATGTTGGCCATGTTTCCCAAGCAAG 1105  
Db 926 GTGGAGGAACCTTTGGAGATTGTTTACTTGTCTCATGTTGGACATGTTTTCGAAAGCTA 985  
QY 1106 CTCCTACTCCCGCAACAGGCTCTGGCCACAGTGT-----TCGTGCAG 1150  
Db 986 CACCTTACAGTTTCCAGGAGGAGCAGGGGAGATTTATCAATAAAATAACAGAGACTTG 1045  
QY 1151 CTGAAGTATGGATGATGAATTTAAAGAGCTCTACTACCATCGCAACCCCGTCCCGCT 1210  
Db 1046 CAGAAGTATGGATGATGAATTTCAAGAAATTTCTTCTATATAATTTCTCCAGGTGTACAA 1105  
QY 1211 TGAACCTTTTGGGGATGTGACAGAGAGAGAGAGCTCCGGGACAAGCTCCAGGTGTAAG 1270  
Db 1106 AGGTAGATTATGGAGATATATCATCAAGACTTGGTCTAAGGCACAACTCCAATGCAGAC 1165  
QY 1271 ACTTCAAGTGGTCTTGGAGACTGTCTATCCAGAACTGCTGCTGAGGAGAGGCGCTG 1330  
Db 1166 CATTCCTTGGTACTAGAGAAATTTTATCTGATTTCTCAGATTCTCGTCACTATTCT 1225  
QY 1331 GCTTCTTCGGGATGCTCCAGAACAAAGAGACTAAACAGACTACTGTTTGACTATAACCTC 1390  
Db 1226 CTTTGGGAGAGATAC-----GAAATGTGGAACAAATCAAGTGTCTAGATAACATGCTA 1279  
QY 1391 CCGATGAAACCAAGATTGTGGGACACCAGGTCAATCTGTACTCTGTATGGGATGGCC 1450

Db 1280 GAAAGAGAAATGAAAAAGTTGGA-----ATTTTAACTGTCTATGGATGGAG 1327  
Qy 1451 AGAATCAGTTTTCAGTACACGTCGCCAGAAAGAAAT 1487  
Db 1328 GTAATCAGGTTTCTCTTACACTGCCCAAGAAAT 1364

RESULT 9  
PCT-US94-02552-18  
; Sequence 18, Application PC/TUS9402552  
; GENERAL INFORMATION:  
; APPLICANT: Elhammer, Ake P.  
; APPLICANT: Homa, Fred L.  
; TITLE OF INVENTION: A Cloned DNA Encoding a UDP-GalNac:  
; TITLE OF INVENTION: Polypeptide, N-Acetylgalactosaminyltransferase  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: The Upjohn Company, Corp. Intellectual  
; ADDRESSEE: Property Law  
; STREET: 301 Henrietta Street  
; CITY: Kalamazoo  
; STATE: Michigan  
; COUNTRY: USA  
; ZIP: 49001  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US94/02552  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Darnley Jr., James D.  
; REGISTRATION NUMBER: 33,673  
; REFERENCE/DOCKET NUMBER: 4755.P CP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 616-385-5210  
; TELEFAX: 616-385-6897  
; TELEX: 224401  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1617 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; PCT-US94-02552-18

Query Match 13.3%; Score 232.2; DB 5; Length 1617;  
Best Local Similarity 54.8%; Pred. No. 5.2e-44;  
Matches 612; Conservative 0; Mismatches 463; Indels 42; Gaps 6;

Qy 392 ATTATGATAATTTGCCAGGACATCTGTATCATAGCAATTTTATAATGAACCTGGTCAA 451  
Db 269 ATCCAGATAACCTTCTCAACACAGTGGTGGTATTTTCCACAAATGAGCTTGGAGCA 328

Qy 452 CTCTCCTTCGGACAGTTTACAGTGTCTTGTAGACATCCCGGATATCCTGTAGAAGAG 511  
Db 329 CACTTCTCGAAGTGTCTCATAGCGTCAATTAATCGCTCAACCAAGGCACATGCTAGAAGAA 388

Qy 512 TGATCCTTGTAGATGACTACAGTGTAGAGACACCTGAAGAGCGCTTGCCCATGAGC 571  
Db 389 TTGTCTGTAGTATGATGCCAGTGAAGAGACTTTTAAAGAGACCTCTAGAGAGTTACG 448

Qy 572 TTTCGGAGT--GCCCAAGTGGCCCTGTATCCCGGCCCAACAGAGAGAGCGCTGGTGC 628  
Db 449 TGAATAAATAAAGTACCGTTACGTCATTCGATGAGCAGCGTTCTGGATTGATCA 508

Qy 629 GAGCCCGCTCTGGGGCGTCTCGCGGAGGGGGCGATGTTCTGACCTTCTGACATGTC 688  
Db 509 CAGCTAGTGTAAAGGTGCTGTGTCTAAAGGCCAAGTATACCTTTTATAGACGGC 568

Qy 689 ACTGTGAGTCCCAAGAGGGTGTCTGAGCGCTGTCTGAGAGGATCCATGAAGAGAGT 748  
Db 569 ACTGTGAGTGCACAGTGGGTGTCTGAGCGCTCTCTTAGCCAGATCAACATGACAGGA 628  
Qy 749 CGGCACTGTGTGTCCTGGGTGATGTGATGACCTGGAAACACTTCTGGAATACCTGGGA 808  
Db 629 AGACAGTGGTCTGTCCCATCATAGATGTGATCAGTGTGATGACACTTTCGAGTACATGSCAG 688  
Qy 809 ACTCCGGGAGCCCAAGATCGCGGTTCGACTGGAGGCTGTGTTCACTGCGACACAG 868  
Db 689 GTTCTGACATGACCTA---TGGCGGTTCAACTGGAAGCTCAACTTTCGCTGGTATCCTG 745  
Qy 869 TTCTCTGAGAGGAGAGG---ATACGGATGCAATCCCGCTCGATGTCTCATGAGTCTCCAA 925  
Db 746 TTCCCAAGAGAAATGGACAGAGGAAGGTGATCGGACTCTTCTGTGGAACACTTA 805  
Qy 926 CAATGGCTGTGTGGCTGTGTGTGAGTAAGAAATATTTTGAATATCTGGGTCTTTATG 985  
Db 806 CAATGGCAGGAGGCTTTTCAATAGACAGAGATTACTTTAGGAAATTTGGAACATATG 865  
Qy 986 ATACAGGAATGGAAGTTTGGGAGGAGAAACCTCGAATTTTCTTTAGGATCTGGCAGT 1045  
Db 866 ATCTGGAATGGATATTTGGGAGGAGAAACCTAGAAATTTCTTTAGGATTTGGCAGT 925  
Qy 1046 GTGTGGGTCTTGGAAACACACCCATGTTCCCATGTGGCCATGTTTCCCAAGCAAG 1105  
Db 926 GTGGAGAACTTTGGAGATGTTTACTTGTCTCATGTGGACATGTTTTCGGAAAGCTA 985  
Qy 1106 CTCCTACTCTCCGCAACAAAGGCTCTGGCCAAACAGTGT-----TCGTGCAG 1150  
Db 986 CACCTTACAGCTTTCCAGGAGGACGCGGCGAGATTATCAATAAAATAACAGACGACTTG 1045  
Qy 1151 CTGAAGTATGGATGGATGAATTTAAAGAGCTCTTACTACCATCGCAACCCCGTCCCGCT 1210  
Db 1046 CAGAAGTATGGATGGATGAATTTCAAGAAATTTCTTATATAATTTCTCCAGGTGTTACAA 1105  
Qy 1211 TGGAACTTTTGGGATGTGACAGAGAGAAAGAGCTCCCGGCAACAGCTCCAGTGTAAAG 1270  
Db 1106 AGGTAGATTATGGAGATATATCATCAAGACTTGTCTAAGGCACAACTCCATGACAG 1165  
Qy 1271 ACTTCAAGTGTCTTGGAGACTGTGTATCCAGAACTGCAATGCTGCTGAGACAGGCGTG 1330  
Db 1166 CATTTCTTGGTACCTAGAGATATTTATCTGATTTCTCAGATTCTCTCGTCACTATTCT 1225  
Qy 1331 GCTTCTCGGATGCTCCAGAAACAAAGGACTTAACAGACTACTGTTTGTACTATAACCTC 1390  
Db 1226 CTTTGGGAGAGATAC-----GAAATGTGGAACAAATCAGTGTCTAGATAACATGGCTA 1279  
Qy 1391 CCAGTAAACACAGATTGTGGGACACCAGGTCAATCTGTACCTCTGTATGGATGGCC 1450  
Db 1280 GAAAAGAGATGAAAAGTTGGA-----ATTTTAACTGTCTATGGATGGAG 1327

Qy 1451 AGAATCAGTTTTCGAGTACACGTCGCCAGAAAGAAAT 1487  
Db 1328 GTAATCAGGTTTCTCTTACACTGCCCAAGAAAT 1364

## RESULT 10

US-08-967-508-11  
; Sequence 11, Application US/08967508  
; Patent No. 5910570

## GENERAL INFORMATION:

; APPLICANT: The Upjohn Company  
; APPLICANT: FOR U.S. PURPOSES ONLY: Elhammer, Ake P. and Homa, Fred L.  
; TITLE OF INVENTION: A Cloned DNA Encoding a UDP-GalNac:  
; TITLE OF INVENTION: Polypeptide, N-Acetylgalactosaminyltransferase  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pharmacia & Upjohn Company, Intellectual  
; ADDRESSEE: Property Legal Services  
; STREET: 301 Henrietta Street  
; CITY: Kalamazoo

STATE: Michigan  
COUNTRY: USA  
ZIP: 49001  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/967,508  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/602,830  
FILING DATE: 13 No. 5910570ember 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Darnley Jr., James D.  
REGISTRATION NUMBER: 33,673  
REFERENCE/DOCKET NUMBER: 4755.P CP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 616-833-2210  
TELEFAX: 616-833-8897  
TELEX: 224401  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1680 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cdna  
US-08-967-508-11

Query Match  
Best Local Similarity 54.8%; Pred. No. 5.3e-44;  
Matches 612; Conservative 0; Mismatches 463; Indels 42; Gaps 6;

392 ATTATCATATTTGCCAGAGATCTGTATCATAGATTTTATATGAAGCTTGTCAA 451  
Db ATCCAGATTAACCTTCTCTACAAACAGTGTGTGTATTTTCCCAATGAGGCTTGGAGCA 391

452 CTCTCTCTCGACAGTTTACAGTGTCTTCCAGACATCCCGGATATCTCTAGAGAAG 511  
Db CACTTCTCGAATCTCCATAGCTCATTAATCGCTCACCAGGCATCTAGAGAAGA 451

512 TGATCTCTGATGACTACAGTATAGAGACACCTGAGAGCGCTTGGCCAAATGAGC 571  
Db TTGTTCTAGTAGATGATGCGAGTGAAGAGACTTTTAAAGACCTCTAGAGAGTTACG 511

572 TTTGCGGACT---GCCAAGTGGCTGTATCCCGCAACAAGAGAGAGCGCTGTGTC 628  
Db TGAATAAATTTAAAGTACCCGTTTACGTCATTCGAATGGAGCAGCGTTCTGGATTGATCA 571

629 GAGCCCGGCTGTGGGGGCTGTCCGCGAGGGGCGATGTTCTGACCTTCTGAGCTGTC 688  
Db GAGCTAGTTAAAGTGTGCTGTCTTAAGCCCAAGTGAATCACTTTTATAGACGCGC 631

689 ACTGTGAGTCCCAAGAGGTGTGAGCGCTGTCTGACAGAGGATCCATGAAGAGAGT 748  
Db ACTGTGAGTGCACAGTGGGGTGTGAGCGCTCTCTAGCCAGATCAACATGACAGGA 691

749 CGGAGTGTGTCGCGGTGATGTATGATGATGATGATGATGATGATGATGATGATGATG 808  
Db AGACAGTGTGTCGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 751

809 ACTCCGGGAGCCCAAGATCGCGGTTTTCGACGTGGAGGCTGTGTCACGTGGCACAG 868  
Db GTTCTGACATGACCTA---TGGCGGGTTCACTGGAAGCTCAACTTTCGCTGTATCCTG 808

869 TTCTCGAGAGGAGAGG---ATACGATGCAATCCCGCTGATGATGATGATGATGATGATG 925  
Db TTCCCAAGAGAGAAATGGACAGAGGAAGGTGATCGGACTCTTCTGCTGAGACACCTA 868

926 CAATGCTGTGGGCTGTGTTGCTGTGAGTAAGAAATATTTTGAATATCTGGGCTTTATG 985

869 CAATGGCAGGAGGCTTTTTCATAGACAGAGATTACTTTAGGAAATTTGSAACATATG 928  
QY ATACAGGAATGGAAGTTTGGGAGGAGAGAAACCTCGAATTTTCTTTAGGATCTGGCAGT 1045  
Db ATGCTGGAATGATATTTGGGAGGAGAGAAACCTAGAAATTTCTTTAGGATTTGGCAGT 988  
QY GTGGTGGGTTCTGGAACACACCCATGTTCCCATGTTGGCCATGTTTCCCAAGCAAG 1105  
Db GTGGAGGAATTTGGAGATTGTTACTTGTCTACATGTTGACATGTTGTTTCGGAAGCTA 1048  
QY CTCCTACTCCCGCAACAGGCTCTGGCCAAAGTGT-----TCGTGCAG 1150  
Db CACCCTACAGTTTCCAGGAGGACCGGGCAGATTATCAATAAAAAATAACAGACGACTTG 1108  
QY CTGAAGTATGGATGATGAATTTAAAGAGCTCTTACTACCATCGAACCCCGCTGCCGCT 1210  
Db CAGAAGTATGGATGATGAATTCNAGAAATTTCTTATATAATTTCTCAGGTGTTACAA 1168  
QY TGAACCTTTTGGGATGTGACAGAGAGGAAGCAGCTCCGGGACAAAGCTCCAGTGTAAAG 1270  
Db AGGTAGATTATGAGATATATCATCAAGACTTGGTCTAAGGCACAAACTCCAATGCAGAC 1228  
QY ACTTCAAGTGGTCTTGGGAGCTGTGTATCCAGAACTGATGCTGAGGACGAGCTG 1330  
Db CATTTCTTGGTACCTAGAGAAATTTATCTCGATTCTCAGATTCTCGTCACTATTCT 1288  
QY GCTTCTTCGGGATGCTCCAGAACAAAGGACTTAACAGACTACTCTTGTGACTATAACCTC 1390  
Db CTTTGGGAGAGATAC-----GAAATGTGGAACAAATCAGTGTCTAGATAACATGGCTA 1342  
QY CGATGAAACACAGATTGTGGGACACACAGGTCAATCTGTACCTCTGTCTGATGGATGGCC 1450  
Db GAAAAGAGAAATGAAAAGTTGGA-----ATTTTAACTGTCTGATGGTATGGAG 1390  
QY AGAATCAGTTTTCAGTACACGTCGCCCAAGAAAT 1487  
Db GTAATCAGGTTTTCTCTTACACTGCCCAACAAAGAAAT 1427

RESULT 11  
US-08-967-506-11  
Sequence 11, Application US/08967506  
Patent No. 6096512  
GENERAL INFORMATION:  
APPLICANT: The Upjohn Company  
APPLICANT: FOR U.S. PURPOSES ONLY: Elhammer, Ake P. and Homa, Fred L.  
TITLE OF INVENTION: A Cloned DNA Encoding a UDP-GalNac:  
TITLE OF INVENTION: Polypeptide, N-Acetylglucosaminyltransferase  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Pharmacia & Upjohn Company, Intellectual  
ADDRESSEE: Property Legal Services  
STREET: 301 Henrietta Street  
CITY: Kalamazoo  
STATE: Michigan  
COUNTRY: USA  
ZIP: 49001  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/967,506  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/602,830  
FILING DATE: 13 No. 6096512ember 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Darnley Jr., James D.  
REGISTRATION NUMBER: 33,673

QY	1211	TGGAACCTTTTGGGATGTGACAGAGAGGAAACGAGCTCCGGGACAAGCTCCACGTGTAAAG	1270
DB	1169	AGGTAGATTATGGAGATATATCATCAAGACTTGGTCTTAAGGCAACAACTCCAATGCAGAC	1328
QY	1271	ACTTCAAGTGGTCTTTGGGAGACTGTCTATCCAGAACTGCATGTGCTGAGGACAGGCCTG	1330
DB	1229	CATTCTCTGGTACCTAGAGATAATTATTCCTGATTCTCAGATTCTCTCGTCACTATTCT	1288
QY	1331	GCCTTCTTCGGGATGCTCCAGAAACAAAGGACTAACAGACTACTGCTTTTGACTATAACCCCTC	1390
DB	1289	CTTTGGGAGAGATAC-----GAAATGTGAAACAAATCAGTGTCTAGATAAACATGGCTA	1342
QY	1391	CGATGAAACACAGATTGGGGACACAGGTCTATTCTGTACCTCTGTCTATGGGATGGCC	1450
DB	1343	GAAGAAGAGATGAAAGAAGTTGGA-----ATTTTAACTGTCTGTTATGGGAG	1390
QY	1451	AGATCAGTCTTTTCGAGTACACGCTCCAGAAAGAAAT	1487
DB	1391	GTAACAGGTTTTCTCTTACACTGCCAACAAAGAAAT	1427
RESULT 12			
PCT-US94-02552-11			
; Sequence 11, Application PC/TUS9402552			
; GENERAL INFORMATION:			
; APPLICANT: Elhammer, Ake P.			
; APPLICANT: Elhammer, Fred L.			
; TITLE OF INVENTION: A Cloned DNA Encoding a UDP-GalNac:			
; TITLE OF INVENTION: Polypeptide, N-Acetylgalactosaminyltransferase			
; NUMBER OF SEQUENCES: 19			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: The Upjohn Company, Corp. Intellectual			
; ADDRESSEE: Property Law			
; STREET: 301 Henrietta Street			
; CITY: Kalamazoo			
; STATE: Michigan			
; COUNTRY: USA			
; ZIP: 49001			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: Floppy disk			
; COMPUTER: IBM PC compatible			
; OPERATING SYSTEM: PC-DOS/MS-DOS			
; SOFTWARE: Patent In Release #1.0, Version #1.25			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: PCT/US94/02552			
; FILING DATE:			
; CLASSIFICATION:			
; ATTORNEY/AGENT INFORMATION:			
; NAME: Darnley Jr., James D.			
; REGISTRATION NUMBER: 33,673			
; REFERENCE/DOCKET NUMBER: 4755.P CP			
; TELECOMMUNICATION INFORMATION:			
; TELEPHONE: 616-385-5210			
; TELEFAX: 616-385-6897			
; TELEX: 224401			
; INFORMATION FOR SEQ ID NO: 11:			
; SEQUENCE CHARACTERISTICS:			
; LENGTH: 1680 base pairs			
; TYPE: nucleic acid			
; STRANDEDNESS: double			
; TOPOLOGY: linear			
; MOLECULE TYPE: cdna			
PCT-US94-02552-11			
Query Match . 13.3%; Score 232.2; DB 5; Length 1680;			
Best Local Similarity 54.8%; Pred. No. 5.3e-44;			
Matches 612; Conservative 0; Mismatches 463; Indels 42; Gaps 6;			
QY	392	ATTATGATAATTTGCCAGGACATCTCTTATCATAGCATTTTATAAGAACCTGGTCAA	451
DB	332	ATCCAGATAACCTTCTTACAAACAGTGGTGATGTTTCCCAATGAGCGCTTGGACCA	391



452	QY	CTCTCTCTCGGACAGTTTACAGTGTCTTGAGACATCCCGGATATCTCTGTAGAAGAAG	511
392	Db	CACCTTCTGCGAACTGTCCATAGGCTCATTTAATCGCTCACCAAGGCACATGCTAGAAGAAA	451
512	QY	TGATCTCTGTAGATGACTACAGTGATAGAGACACCTGAAGGAGCGCTTCGCCAATGAC	571
452	Db	TTGTTCTAGTAGATGATGCCAGTGAAGAGACTTTTTAAAGAAGACCTCTAGAGAGTTACG	511
572	QY	TTTTCGGAGCT---GCCAAGGTGCGCTGATCCGCGCAACAAGAGAGAGGGCCCTGGTGC	628
512	Db	TGAAAAAATAAAGTAGTACCCTGTTACGCTCATTCGAATGGAGCAGCTTCTGGATTGATCA	571
629	QY	GAGCCCGGCTGCTGGGGCGTCTGCGCGAGGGCGATGTTCTGAACCTTCTCTGACCTGTCT	688
572	Db	GAGCTAGGTTTAAAGGTGCTGCTGTGCTTAAAGGCCAAGTGATCACCTTTTTTAGACGCGC	631
689	QY	ACTGTAGTCCACGAAGGGTGGCTGAGCGCTGCTGCAGAGGATCCATGAAGAGAGT	748
632	Db	ACTGTAGTGCAAGTGGGGTGGCTGGAGCCTCTCTTAGCCAGGATCAAAACATGACAGGA	691
749	QY	CGCGAGTGGTGTGCCCGGTGATTGATGTGATCGACTGGAAACACCTTCGAATACCTGGCGGA	808
692	Db	AGACAGTGGTCTGTCCCATCATAGATGTGATCAGTGATGACACTTTCGAGTACATGCGAG	751
809	QY	ACTCGGGGAGCCACAGATCGCGGTTTCACTGGAGGCTGGTGTTCACGTGGCACACAG	868
752	Db	GTTCGACATGACCTA---TGGCGGGTTCAACTGGAAAGCTCAACTTTTCGCTGGTATCCTG	808
869	QY	TTCTGAGAGGGAGAGG---ATACGGATGCAATCCCCCGTGGATGTGATCAGGTCTCCAA	925
809	Db	TTCCCCAAGAGAGAAATGGACAGAGGAAGGTGATCGGACTCTTCTCTGTGAGAAACACTTA	868
926	QY	CAATGGCTGGTGGGCTGTTGCTGTGATGAAGAAATATTTGATATCTCGGGTCTTTATG	985
869	Db	CAATGGCAGAGGCGCTTTTTTCAATGACAGAGATTACTTTCAGGAAATTTGGAACATATG	928
986	QY	ATACAGGAATGGAAATTTGGGGAGGAGAAACCTCGAAATTTCTCTTTAGGATCTGGCAGT	1045
929	Db	ATGCTGGAATGGATATTTGGGGAGGAGAAACCTAGAAAATTTCTTTAGGATTTGGCAGT	988
1046	QY	GTGGTGGGTTTCTGGAACAACACCCATGTTCCCATGTTGGCCATGTTTTCCCAAGCAAG	1105
989	Db	GTGGAGGAACTTTGGAGATTGTTACTTGTCTCAGATGTGGACATGTGTTTCGGAAAGCTA	1048
1106	QY	CTCCCTACTCCCGCAACAGGCTCTGCCAACAGTGT-----TCGTGCGAG	1150
1049	Db	CACCTTACAGTTTCCAGGAGGACCGGGGAGATTTATCAATAAATAAATPAAACAGACACTTG	1108
1151	QY	CTGAAGTATGGATGGATGAATTTAAAGAGCTCTACTACCATTGCAACCCCGCTGCCCGCT	1210
1109	Db	CAGAAGTATGGATGGATGAATTCAGAAATTTCTTCTATATAATTTCTCCAGGTGTTACAA	1168
1211	QY	TGAAACCTTTTGGGATGTGACAGAGAGGAGCAGCTCCGGGACAAGCTCCAGTGTAAAG	1270
1169	Db	AGGTAGATTATGAGATATATCATCAAGACTTGGTCTTAAGGCACAAACTCCAATGACAG	1228
1271	QY	ACTTCAAGTGGTCTTGGAGACTGTGTATCCAGAACTGCAATGTCCTGAGGACAGGCGCTG	1330
1229	Db	CATTCTCTTGGTACCTAGAGAAATTTATCCTGATTTCTCAGATTCTCTCGTCACTATTTC	1388
1331	QY	GCCTTCTCGGGATGCTCCAGAAACAAGGACTAACAGACTACTGCTTTTGACTATAACCTC	1390
1289	Db	CTTTGGAGAGATAC-----GAAATGTGAAACAATACTAGTGTCTAGATAACATGGCTA	1342
1391	QY	CCGATGAAACACAGATTGTGGGACACACAGGTCAATCTGTACCTCTGTATGGGATGGGCC	1450
1343	Db	GAAAGAGAGATGAAGAAGTTGGA-----ATTTTAACTGTCTGATATGGAG	1390
1451	QY	AGAACTAGTTTTTCAGTACACGTCCCGAAGAGAAAT	1487
1391	Db	GTAATCAGGTTTTTCTCTTACACTGCCAACCAAGAAAT	1427

```

RESULT 13
US-08-967-508-10
; Sequence 10, Application US/08967508
; Patent No. 5910570
; GENERAL INFORMATION:
; APPLICANT: The Upjohn Company
; APPLICANT: FOR U.S. PURPOSES ONLY: Elhammer, Ake P. and Homa, Fred L.
; TITLE OF INVENTION: A Cloned DNA Encoding a UDP-GalNac:
; TITLE OF INVENTION: Polypeptide, N-Acetylgalactosaminyltransferase
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pharmacia & Upjohn Company, Intellectual
; ADDRESSEE: Property Legal Services
; STREET: 301 Henrietta Street
; CITY: Kalamazoo
; STATE: Michigan
; COUNTRY: USA
; ZIP: 49001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/967,508
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/602,830
; FILING DATE: 13 No. 5910570ember 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Darnley Jr., James D.
; REGISTRATION NUMBER: 33,673
; REFERENCE/DOCKET NUMBER: 4755.P CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 616-833-2210
; TELEFAX: 616-833-8897
; TELEX: 224401
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2294 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-967-508-10

Query Match 13.3%; Score 232.2; DB 2; Length 2294;
Best Local Similarity 54.8%; Pred. No. 5.9e-44;
Matches 612; Conservative 0; Mismatches 463; Indels 42; Gaps 6;

QY 392 ATTATGATAATTGCCAGGACATCTGTATCATAGCATTTTATTAATGAAGCCCTGGTCAA 451
Db 384 ATCCAGATAACCTTCCTACACCAGTGTGGTATGTTTTCACAATGAGGCTTGGAGCA 443
QY 452 CTCCTCCTTCGGACAGTTTTACAGTGTCCTTGAGACATCCCGGATATCCTGCTAGAAGAAG 511
Db 444 CACTTCTGGAACTGTCCATAGCGTCATTATCGCTACCAAGGCACATGCTAGAGAAA 503
QY 512 TGATCCTTGTAGATGACTACAGTGATAGAGAGCACCTGGAAGAGCGCTTGCCCAATGAGC 571
Db 504 TTGTTCTAGTAGATGATGCCAGTGAAGAGACTTTTTTAAAAAGACCTCTAGAGAGTTACG 563
QY 572 TTTCCGGACT--GCCCAAGGTGCGCTGATCCGGCCCAACAAGAGAGAGGCGCTGGTGC 628
Db 564 TGAAAAAATAAAAGTAGACCCGTTACGTCATTCGAATGGAGACGCGTTCTGGATTGATCA 623
QY 629 GAGCCCGCGTCTGTGGGGCGGTCTGCGGCGGGGGCGATGTTCTGACCTTCCTGGACTGTC 688
Db 624 GAGCTAGGTTAAAGGTGCTGCTGTGTCCTAAGGCCAAGTCAACCTTTTACACGGCG 683
QY 689 ACTGTGAGTGCCACGAAGGGTGGCTGGAGCCGCTGCTGACAGAGGATCCATGAAGAGGAGT 748

```





Db 1281 CATTCTCTTGGTACCTAGAGAATATTTATCCTGATTTCTCAGATTCTCGTCACCTATTTTCT 1340  
Qy 1331 GCTTCTTCGGGATGCTCCAGAACAAAGGACTAACAGACTACTGCTTTGACTATATAACCCCTC 1390  
Db 1341 CTTTGGGAGAGATAC-----GAAATGTGGAAACAAATCAGTGTCTAGATAACATGGCTA 1394  
Qy 1391 CCGATGAAACCAGATTGTGGGACACCAGGTCAATTCTGTACCTCTGTCAATGGGATGGGCC 1450  
Db 1395 GAAAGAGATGAAAGTTGGA-----ATTTTAACTGTCAATGGTATGGGAG 1442  
Qy 1451 AGAATCAGTTTTTCGAGTACACGTCGCCAGAAAGAAAT 1487  
Db 1443 GTAATCAGTTTTTCTTTACACTGCCAACAAAGAAAT 1479

Search completed: December 7, 2003, 21:21:39  
Job time : 109.153 secs



```
Qy 121 CGTGGGCGGGCGGGCTGCGAGCCGGGACCCCGCGCACCCCGCGCGCGCGCGG 180
Db 121 CGTGGGCGGGCGGGCTGCGAGCCGGGACCCCGCGCACCCCGCGCGCGCGG 180
Qy 181 GCGAGCGGTCATGCGCGCGCGCGCGGTCGCGCGGAAACGCGCTGGCGCGCGGGCGAG 240
Db 181 GCGAGCGGTCATGCGCGCGCGCGCGGTCGCGCGGAAACGCGCTGGCGCGCGGGCGAG 240
Qy 241 GCGGTGGGCTGCGAGCTGCGAGGCGAGAGCTGCGGCTGCGAGGAGAGCGTGGCGCTG 300
Db 241 GCGGTGGGCTGCGAGCTGCGAGGCGAGAGCTGCGGCTGCGAGGAGAGCGTGGCGCTG 300
Qy 301 CACGAGATTACATCTACCTCAGCGACCGCATCTCACTGCACCCGCGCTGCCCGAGCGC 360
Db 301 CACGAGATTACATCTACCTCAGCGACCGCATCTCACTGCACCCGCGCTGCCCGAGCGC 360
Qy 361 TGGAAACCGCTGTGCAAGAGAGAAATATGATTTATGATTAATTTGCCAGACATCTGTT 420
Db 361 TGGAAACCGCTGTGCAAGAGAGAAATATGATTTATGATTAATTTGCCAGACATCTGTT 420
Qy 421 ATCATAGCATTTTATATGAGCTGCTCACTCTCTCTCGGACAGTTTACAGTGTCTTT 480
Db 421 ATCATAGCATTTTATATGAGCTGCTCACTCTCTCTCGGACAGTTTACAGTGTCTTT 480
Qy 481 GAGACATCCCGGATATCCTCTAGAGAGCTGATCTCTTTGATGATGACTACAGTGATAGA 540
Db 481 GAGACATCCCGGATATCCTCTAGAGAGCTGATCTCTTTGATGATGACTACAGTGATAGA 540
Qy 541 GAGCACTGAAAGGAGCGCTTGGCCAAATGAGCTTTTGGGACTGCCCCAAGGTGCGCTGATC 600
Db 541 GAGCACTGAAAGGAGCGCTTGGCCAAATGAGCTTTTGGGACTGCCCCAAGGTGCGCTGATC 600
Qy 601 CGCGCAACAGAGAGAGGCGCTGGTCGAGACCGGCTGCTGGGGCGCTGCGCGCGAGG 660
Db 601 CGCGCAACAGAGAGAGGCGCTGGTCGAGACCGGCTGCTGGGGCGCTGCGCGCGAGG 660
Qy 661 GCGGATGTTCTGACCTTCTGAGCTGTCTGAGTGTGAGTGCCACGAAGGTGCTGGAGCG 720
Db 661 GCGGATGTTCTGAGCTTCTGAGCTGTCTGAGTGTGAGTGCCACGAAGGTGCTGGAGCG 720
Qy 721 CTGCTGACAGAGATCCATGAAGAGAGTTCGCGAGTGTGTCGCCGCTGATTTGATGTGATC 780
Db 721 CTGCTGACAGAGATCCATGAAGAGAGTTCGCGAGTGTGTCGCCGCTGATTTGATGTGATC 780
Qy 781 GACTGGACACTTCGATACCTCGGNACTCCGCGGAGCCCGAGTCCGCGGTTTCGAC 840
Db 781 GACTGGACACTTCGATACCTCGGNACTCCGCGGAGCCCGAGTCCGCGGTTTCGAC 840
Qy 841 TGGAGGCTGGTGTTCACGTGGCACACAGTTCCTGAGAGGAGAGATACGATGCAATCC 900
Db 841 TGGAGGCTGGTGTTCACGTGGCACACAGTTCCTGAGAGGAGAGATACGATGCAATCC 900
Qy 901 CCGGTCGATGTATCAGGTCTCCAAATGCTGCTGGGCTGTTTGTGTGAGTAAGAAA 960
Db 901 CCGGTCGATGTATCAGGTCTCCAAATGCTGCTGGGCTGTTTGTGTGAGTAAGAAA 960
Qy 961 TATTTGATATCTGGGCTCTATGATACAGGAATGGAAGTTTGGGAGGAGAAACCTC 1020
Db 961 TATTTGATATCTGGGCTCTATGATACAGGAATGGAAGTTTGGGAGGAGAAACCTC 1020
Qy 1021 GAAATTTCTTTAGGATCTGCGAGTGTGGTGGGTTCTGGAAACACACCCATGTTCCCAT 1080
Db 1021 GAAATTTCTTTAGGATCTGCGAGTGTGGTGGGTTCTGGAAACACACCCATGTTCCCAT 1080
Qy 1081 GTTGCCCATGTTTCCCAAGCAAGCTCCCTACTCCCGCAACAAAGGCTCTGGCCAAACAGT 1140
Db 1081 GTTGCCCATGTTTCCCAAGCAAGCTCCCTACTCCCGCAACAAAGGCTCTGGCCAAACAGT 1140
Qy 1141 GTTGTGCGAGCTGAAGTATGATGATGAATTAAGAGCTCTACTACCATCGCAACCCC 1200
Db 1141 GTTGTGCGAGCTGAAGTATGATGATGAATTAAGAGCTCTACTACCATCGCAACCCC 1200
Qy 1201 CGTGGCGCTTGGAAACCTTTTGGGATGTGACAGAGAGAGAGCTCCGGGACAAAGCTC 1260
```

```
Db 1201 CGTGGCGCTTGGAAACCTTTTGGGATGTGACAGAGAGAGAGCTCCGGGACAAAGCTC 1260
Qy 1261 CAGTGTAAAGACTTCAAGTGGTCTTGGAGACTGTGTATCCAGAACTGCACTGTGCCTGAG 1320
Db 1261 CAGTGTAAAGACTTCAAGTGGTCTTGGAGACTGTGTATCCAGAACTGCACTGTGCCTGAG 1320
Qy 1321 GACAGGCTGGCTTCTTGGGATGTCTCCAGAAACAAAGGACTAAACAGACTACTGCTTTGAC 1380
Db 1321 GACAGGCTGGCTTCTTGGGATGTCTCCAGAAACAAAGGACTAAACAGACTACTGCTTTGAC 1380
Qy 1381 TATAACCTCCCGATGAAACACAGATTGTGGGACACAGGTCACTTGTGTACCTGTGTCAT 1440
Db 1381 TATAACCTCCCGATGAAACACAGATTGTGGGACACAGGTCACTTGTGTACCTGTGTCAT 1440
Qy 1441 GGGATGGGCGAGAAATCAGTTTTTTCGAGTACACGTCCTCCAGAAAGAAATACGCTATAACACC 1500
Db 1441 GGGATGGGCGAGAAATCAGTTTTTTCGAGTACACGTCCTCCAGAAAGAAATACGCTATAACACC 1500
Qy 1501 CACGAGCTGAGGCTGTCATTGTGTGGAAGCAGGAATGATACCTTTATCATGTCATCTC 1560
Db 1501 CACGAGCTGAGGCTGTCATTGTGTGGAAGCAGGAATGATACCTTTATCATGTCATCTC 1560
Qy 1561 TGGCAAGAAACTCCCCAGAGAAATCAGAACTTCACTTTCAGGAGGATGGAATCTTTATTT 1620
Db 1561 TGGCAAGAAACTCCCCAGAGAAATCAGAACTTCACTTTCAGGAGGATGGAATCTTTATTT 1620
Qy 1621 CAGCAAGCTCCAGAAATGTGTCCAGGCTGCGAGGAGAGTTCGAGTGCAGTTCGTT 1680
Db 1621 CAGCAAGCTCCAGAAATGTGTCCAGGCTGCGAGGAGAGTTCGAGTGCAGTTCGTT 1680
Qy 1681 CCACTCTTACGAGACTGCACCAACTCGGATCATCAGAAATGGTTCTTCAAAGAGCGCATG 1740
Db 1681 CCACTCTTACGAGACTGCACCAACTCGGATCATCAGAAATGGTTCTTCAAAGAGCGCATG 1740
Qy 1741 TTATGA 1746
Db 1741 TTATGA 1746
```

## RESULT 2

```
US-10-074-527-1
; Sequence 1, Application US/10074527
; Publication No. US20020142426A1
; GENERAL INFORMATION:
; APPLICANT: Olandt, Peter J.
; APPLICANT: Meyers, Rachel E.
; APPLICANT: Galvin, Katherine A.
; APPLICANT: Millennium Pharmaceuticals Inc.
; TITLE OF INVENTION: 33945, A Human Glycosyltransferase and
; FILE OF INVENTION: Uses Therefor
; FILE REFERENCE: MPI2001-018P1RCP1(M)
; CURRENT APPLICATION NUMBER: US/10/074,527
; CURRENT FILING DATE: 2002-02-12
; PRIOR APPLICATION NUMBER: 60/269202
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2850
; TYPE: DNA
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (81)...(1826)
US-10-074-527-1
```

Query Match 100.0%; Score 1746; DB 13; Length 2850;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1746; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 1 ATGTGGGGGCGCACGGCGCGCGCGCTGCTCCCGGAACTGCGCGCGCGCGGAGCG 60
|||||
```

Db 81 ATGTGGGGCGCAGCGGCGGCGCTGCCCCGGGAACTGCGGCGCGCGCGGAGCG 140  
Qy 61 CTGTTGTGTCCTGGCGCTACTGGCGTTGGCGGGCTGGGCTGGGCTGGCGGCGAG 120  
Db 141 CTGTTGTGTCCTGGCGCTACTGGCGTTGGCGGGCTGGGCTGGGCTGGCGGCGAG 200  
Qy 121 CGTGGGGCGGGGCGGGGCTGGCGAGCGGGACCCCGCGCACCCCGCGCCCCGGGCGG 180  
Db 201 CGTGGGGCGGGGCGGGGCTGGCGAGCGGGACCCCGCGCACCCCGCGCCCCGGGCGG 260  
Qy 181 CGGAGCGCGTCTATGCGCGCGCGCGCTGGCGGAAACGCTGGGCGCGCGGGGCGAG 240  
Db 261 CGGAGCGCGTCTATGCGCGCGCGCGCTGGCGGAAACGCTGGGCGCGCGGGGCGAG 320  
Qy 241 GCGGTGCGGTGTCAGCTGCGAGGGCGAGAGCTGCGGCTGCGAGGAGAGAGCTGCGGCTG 300  
Db 321 GCGGTGCGGTGTCAGCTGCGAGGGCGAGAGCTGCGGCTGCGAGGAGAGAGCTGCGGCTG 380  
Qy 301 CACGAGATTAACTACTACTCAGGACCGCATCTCACTGACCGCGCGCTGCCGAGCGC 360  
Db 381 CACGAGATTAACTACTACTCAGGACCGCATCTCACTGACCGCGCGCTGCCGAGCGC 440  
Qy 361 TGGAAACCGCTGTGCAAGAGAGAAATATGATTATGATAATTGGCCAGGACATCTGTT 420  
Db 441 TGGAAACCGCTGTGCAAGAGAGAAATATGATTATGATAATTGGCCAGGACATCTGTT 500  
Qy 421 ATCATAGCATTTTATATAGAGCTGTGCAACTCTCTTCGAGAGTTTACAGTGTCTTT 480  
Db 501 ATCATAGCATTTTATATAGAGCTGTGCAACTCTCTTCGAGAGTTTACAGTGTCTTT 560  
Qy 481 GAGACATCCCGGATATCTCTGCTAGAGAGAGTATCTTGTAGATGACTACAGTGATAGA 540  
Db 561 GAGACATCCCGGATATCTCTGCTAGAGAGAGTATCTTGTAGATGACTACAGTGATAGA 620  
Qy 541 GAGCACCCTGAAGAGAGCGCTTGGCCAAATGAGCTTTCGGGACTGCCCAAGGTCGCTGATC 600  
Db 621 GAGCACCCTGAAGAGAGCGCTTGGCCAAATGAGCTTTCGGGACTGCCCAAGGTCGCTGATC 680  
Qy 601 CGCGCCAAAGAGAGAGGCGCTGTGTCGAGCGCGCTGTGCGGGCGTGTGCGGCGAGG 660  
Db 681 CGCGCCAAAGAGAGAGGCGCTGTGTCGAGCGCGCTGTGCGGGCGTGTGCGGCGAGG 740  
Qy 661 GCGGATGTTCTGACCTTCTGAGTGTGCTGAGTGTGCTGAGTGTGCTGAGTGTGCTGAG 720  
Db 741 GCGGATGTTCTGACCTTCTGAGTGTGCTGAGTGTGCTGAGTGTGCTGAGTGTGCTGAG 800  
Qy 721 CTGCTGAGAGGATCCATGAAGAGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGG 780  
Db 801 CTGCTGAGAGGATCCATGAAGAGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGG 860  
Qy 781 GACTGGAACACCTTTCGAATACCTGGGAACTCCGGGAGCGCCAGATCGGCGGTTTCGAC 840  
Db 861 GACTGGAACACCTTTCGAATACCTGGGAACTCCGGGAGCGCCAGATCGGCGGTTTCGAC 920  
Qy 841 TGAGGCTGTGTTTACGTGGGACACAGTTCCTGAGAGGAGAGGATACGGATGCAATCC 900  
Db 921 TGAGGCTGTGTTTACGTGGGACACAGTTCCTGAGAGGAGAGGATACGGATGCAATCC 980  
Qy 901 CCGGTGATGTCATCAGTCTCCAACTGCTGGGCTGTGGGCTGTGGGCTGTGGGCTGTGGG 960  
Db 981 CCGGTGATGTCATCAGTCTCCAACTGCTGGGCTGTGGGCTGTGGGCTGTGGGCTGTGGG 1040  
Qy 961 TATTTTGAATATCTGGGCTTATGATACAGGAATGGAAGTTTGGGAGGAGAAACCTC 1020  
Db 1041 TATTTTGAATATCTGGGCTTATGATACAGGAATGGAAGTTTGGGAGGAGAAACCTC 1100  
Qy 1021 GAATTTTCTTTAGGATCTGGGCTGTGGGCTTCTGAAACACACCCATGTTCCCAT 1080  
Db 1101 GAATTTTCTTTAGGATCTGGGCTGTGGGCTTCTGAAACACACCCATGTTCCCAT 1160  
Qy 1081 GTTGGCCATGTTTTCCCGAGAGCTCCCTACTCCCGGCAACAGGCTCTGGCCACAGT 1140  
Db 1161 GTTGGCCATGTTTTCCCGAGAGCTCCCTACTCCCGGCAACAGGCTCTGGCCACAGT 1220

## RESULT 3

US-10-198-846-9749/c

; Sequence 9749, Application US/10198846

; Publication No. US20030099974A1

; GENERAL INFORMATION:

; APPLICANT: Lillie, James

; APPLICANT: Xu, Yongyao

; APPLICANT: Wang, Youzhen

; APPLICANT: Steinmann, Kathleen

; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS

; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND

; TITLE OF INVENTION: THERAPY OF BREAST CANCER

; FILE REFERENCE: MRI-049

; CURRENT APPLICATION NUMBER: US/10/198,846

; PRIOR FILING DATE: 2002-07-18

; PRIOR APPLICATION NUMBER: 60/306,220

; PRIOR FILING DATE: 2001-07-18

; NUMBER OF SEQ ID NOS: 14084

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 9749

; LENGTH: 2904

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: 1, 2, 3, 4, 5, 2903, 2904

; OTHER INFORMATION: n = A,T,C or G

US-10-198-846-9749

Query Match 100.0%; Score 1746; DB 14; Length 2904;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1746; Conservative 0; Mismatches 0; Indels 0;

Qy	1	ATGTGGGGCGCACGGGCGGGCGGCTGCCCCGCGGAACTGCGGCGCGCGCGGAGCGGCG	60
Db	2826	ATGTGGGGGGCGACAGGCGGCGGGCGGTGCCCGCGGAACTGCGGCGCGCGCGCGGAGCGG	2767
Qy	61	CTGTGGTGTCTCTTGGCGCTACTTGGCGTGTGGCGGSGCTGGCTCGTGTGTGGGGCGCAG	120
Db	2766	CTGTGTGTGTCTCTTGGCGCTACTTGGCGTGTGGCGGSGCTGGCTCGTGTGTGGGGCGCAG	2707
Qy	121	CGTGGGCGCGGGGCGGGGCTGCCGAGCCGGGACCCCGCGCAACCCCGCGCCCGCGCGCG	180
Db	2706	CGTGGGCGCGGGGCGGGGCTGCCGAGCCGGGACCCCGCGCAACCCCGCGCCCGCGCGCG	2647
Qy	181	CGGAGCCGGTTCATGCCCGGGCGCGGTGCCCGGGGAAACGCGCTGGGCGCGCGGGCGCAG	240
Db	2646	CGGAGCCGGTTCATGCCCGGGCGCGGTGCCCGGGGAAACGCGCTGGGCGCGCGGGCGCAG	2587
Qy	241	GCGTGGCGGCTGCAGCTGCAGGCGAGGAGCTCGGCTGCAGGAGGAGAGCGTGGCGCTG	300
Db	2586	GCGTGGCGGCTGCAGCTGCAGGCGGAGGAGCTCGGCTGCAGGAGGAGAGCGTGGCGCTG	2527
Qy	301	CACCAGATTAACTCTACCTCAGCGACCGCATCTCACTGCACCGCCCTGCCGCCGAGCGC	360
Db	2526	CACCAGATTAACTCTACCTCAGCGACCGCATCTCACTGCACCGCCCTGCCGCCGAGCGC	2467
Qy	361	TGGNACCGCTGTGCAAGAGAGAAATATGATTATGATTAATTTGCCAGGACATCTGTT	420
Db	2466	TGGNACCGCTGTGCAAGAGAGAAATATGATTATGATTAATTTGCCAGGACATCTGTT	2407
Qy	421	ATCATAGCATTTTATAATGAAGCTGTGCTCAACTCTCTCTCGGACAGTTTACAGTGTCTT	480
Db	2406	ATCATAGCATTTTATAATGAAGCTGTGCTCAACTCTCTCTCGGACAGTTTACAGTGTCTT	2347
Qy	481	GAGACATCCCGGATATCTGTCTAGAGAAAGTAGTGAATCTTGTAGATGACTACAGTGATAGA	540
Db	2346	GAGACATCCCGGATATCTGTCTAGAGAAAGTAGTGAATCTTGTAGATGACTACAGTGATAGA	2287
Qy	541	GAGCACCTGAGGAGCGCTTGGCCAAATGAGCTTTCGGGACTGCCCAAGGTGCGCCTGATC	600
Db	2286	GAGCACCTGAGGAGCGCTTGGCCAAATGAGCTTTCGGGACTGCCCAAGGTGCGCCTGATC	2227
Qy	601	CGCGCCCAACAGAGAGAGGCGCTGGTCCGAGCCCGGCTGCTGGGGGCGTCTCGCGCGAGG	660
Db	2226	CGCGCCCAACAGAGAGAGGCGCTGGTCCGAGCCCGGCTGCTGGGGGCGTCTCGCGCGAGG	2167
Qy	661	GGCGATGTTCTGACCTTCTTGGAATGCTGAGTGCCAGAGGTCGCCAGAGGTGGCTGGAGCCG	720
Db	2166	GGCGATGTTCTGACCTTCTTGGAATGCTGAGTGCCAGAGGTCGCCAGAGGTGGCTGGAGCCG	2107
Qy	721	CTGCTGCAGAGGATCCATGAGAGGAGTGCAGAGTGTGTGTCGGTGAATTTGATGTGATC	780
Db	2106	CTGCTGCAGAGGATCCATGAGAGGAGTGCAGAGTGTGTGTCGGTGAATTTGATGTGATC	2047
Qy	781	GACTTGGAAACACTTTCGAATACCTCGGGGAACTCCGGGGAGCCCCAGATCGCGGTTTTCGAC	840
Db	2046	GACTTGGAAACACTTTCGAATACCTCGGGGAACTCCGGGGAGCCCCAGATCGCGGTTTTCGAC	1987
Qy	841	TGGAGGCTGTGTTCAGCTGGGCAACAGTTTCTGAGAGGGAGAGGATACGGATGCAATCC	900
Db	1986	TGGAGGCTGTGTGTTCAGCTGGGCAACAGTTTCTGAGAGGGAGAGGATACGGATGCAATCC	1927
Qy	901	CCCGTTCGATGTCATCAGCTGTCCAAACAAATGCGTGGTGGGCTGTTTGTCTGTGAGTAAGAAA	960
Db	1926	CCCGTTCGATGTCATCAGCTGTCCAAACAAATGCGTGGTGGGCTGTTTGTCTGTGAGTAAGAAA	1867
Qy	961	TATTTTGAATATCTGGGGTCTTATGATACAGGAAATGGAAGTTTGGGGAGAGAAAACTTC	1020
Db	1866	TATTTTGAATATCTGGGGTCTTATGATACAGGAAATGGAAGTTTGGGGAGAGAAAACTTC	1807









## RESULT 8

US-09-895-298-44  
; Sequence 44, Application US/09895298  
; Publication No. US20030078405A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: 47 Human Secreted Proteins  
; FILE REFERENCE: P2035P1  
; CURRENT APPLICATION NUMBER: US/09/895,298  
; PRIORITY FILING DATE: 2001-07-02  
; PRIOR APPLICATION NUMBER: 09/591,16  
; PRIOR FILING DATE: 2000-06-09  
; PRIOR APPLICATION NUMBER: PCT/US99/29950  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: 60/113,006  
; PRIOR FILING DATE: 1998-12-18  
; PRIOR APPLICATION NUMBER: 60/112,809  
; PRIOR FILING DATE: 1998-12-17  
; NUMBER OF SEQ ID NOS: 231  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 44  
; LENGTH: 1517  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (144)  
; OTHER INFORMATION: n equals a,t,g, or c  
US-09-895-298-44

Query Match 40.8%; Score 711.6; DB 11; Length 1517;  
Best Local Similarity 99.9%; Pred. No. 2.1e-189;  
Matches 711; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1035 GATCTGGCAGTGTGGTGGGTTCTGGAAACACACCCATGTTCCCATGTTGCCCATGTTTT 1094  
Db 479 GATCTGGCAGTGTGGTGGGTTCTGGAAACACACCCATGTTCCCATGTTGCCCATGTTTT 538  
  
Qy 1095 CCCCAGCAAGCTCCCTACTCCCGCAACAGGCTCTGCCCAACAGTGTTCGTGCAGCTGA 1154  
Db 539 CCCCAGCAAGCTCCCTACTCCCGCAACAGGCTCTGCCCAACAGTGTTCGTGCAGCTGA 598  
  
Qy 1155 AGTATGGATGGATGAATTTAAAGAGCTCTACTACCATCGCAACCCCGTCCCGCTTGGGA 1214  
Db 599 AGTATGGATGGATGAATTTAAAGAGCTCTACTACCATCGCAACCCCGTCCCGCTTGGGA 658  
  
Qy 1215 ACCTTTTGGGATGTGACAGAGAGAAAGCAGCTCCGGGACAAAGCTCCAGTGTAAAGACTT 1274  
Db 659 ACCTTTTGGGATGTGACAGAGAGAAAGCAGCTCCGGGACAAAGCTCCAGTGTAAAGACTT 718  
  
Qy 1275 CAAGTGGTCTTGGAGACTGTGTATCCAGAACTGCATGTGCTGAGGACAGGCTTGCTT 1334  
Db 719 CAAGTGGTCTTGGAGACTGTGTATCCAGAACTGCATGTGCTGAGGACAGGCTTGCTT 778  
  
Qy 1335 CTTCCGGGATGCTCCAGAAACAAAGGACTAAACAGACTTCTGTCTTGAATAACCTCCCGA 1394  
Db 779 CTTCCGGGATGCTCCAGAAACAAAGGACTAAACAGACTTCTGTCTTGAATAACCTCCCGA 838  
  
Qy 1395 TGAACACAGATGTGGGACACCAAGTCAATCTGTACTCTGTCTGTATGGATGGGCGCAGAA 1454  
Db 839 TGAACACAGATGTGGGACACCAAGTCAATCTGTACTCTGTCTGTATGGATGGGCGCAGAA 898  
  
Qy 1455 TCAGTTTTTCAGTACAGCTCCAGAAAGAAATAGCTTATACACCCACAGGCTGAGGG 1514  
Db 899 TCAGTTTTTCAGTACAGCTCCAGAAAGAAATAGCTTATACACCCACAGGCTGAGGG 958  
  
Qy 1515 CTGATTTCTGTGGAGCAGGAATGGATACCTTTATCATGCTCTCTCGGAAGAACTGTC 1574  
Db 959 CTGATTTCTGTGGAGCAGGAATGGATACCTTTATCATGCTCTCTCGGAAGAACTGTC 1018  
  
Qy 1575 CCCAGAGAAATCAGAAATTTCACTTTGACAGAGATGATCTTTATTTACGAACAGTCCAA 1634  
Db 1019 CCCAGAGAAATCAGAAATTTCACTTTGACAGAGATGATCTTTATTTACGAACAGTCCAA 1078

Qy 1635 GAAATGTGTCCAGGCTCCGAGGAAGAGTCCAGTGCAGTTCGTTCCACTCTTTACGAGA 1694  
Db 1079 GAAATGTGTCCAGGCTCCGAGGAAGAGTCCAGTGCAGTTCGTTCCACTCTTTACGAGA 1138  
  
Qy 1695 CTGCACCAACTCGGATCATCAGAAATGGTTCTTCAAAGAGCGCATGTTATGA 1746  
Db 1139 CTGCACCAACTCGGATCATCAGAAATGGTTCTTCAAAGAGCGCATGTTATGA 1190

## RESULT 9

US-10-292-896-121  
; Sequence 121, Application US/10292896  
; Publication No. US20030186850A1  
; GENERAL INFORMATION:  
; APPLICANT: HASSAN, Helle  
; APPLICANT: REIS, Celso A.  
; APPLICANT: BENNETT, Eric P.  
; APPLICANT: CLAUSEN, Henrik  
; TITLE OF INVENTION: METHODS TO IDENTIFY AGENTS MODULATING FUNCTIONS OF POLYPEPTIDE GA  
; TITLE OF INVENTION: TRANSFERASES, PHARMACEUTICAL COMPOSITIONS COMPRISING SUCH AGENTS  
; TITLE OF INVENTION: SUCH AGENTS FOR PREPARING MEDICAMENTS  
; FILE REFERENCE: 4305/1H154-US3  
; CURRENT APPLICATION NUMBER: US/10/292,896  
; PRIORITY FILING DATE: 2003-03-17  
; PRIOR APPLICATION NUMBER: US 60/425,204  
; PRIOR FILING DATE: 2002-11-08  
; PRIOR APPLICATION NUMBER: PCT/DK01/00328  
; PRIOR FILING DATE: 2001-05-10  
; PRIOR APPLICATION NUMBER: US 60/203,331  
; PRIOR FILING DATE: 2000-05-11  
; NUMBER OF SEQ ID NOS: 130  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 121  
; LENGTH: 525  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-292-896-121

Query Match 30.0%; Score 524; DB 12; Length 525;  
Best Local Similarity 100.0%; Pred. No. 7.3e-137;  
Matches 524; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1223 GGGATGTGACAGAGAGGAAGCAGCTCCGGGACAAAGCTCCAGTGTAAAGACTTCAAGTGT 1282  
Db 2 GGGATGTGACAGAGAGGAAGCAGCTCCGGGACAAAGCTCCAGTGTAAAGACTTCAAGTGT 61  
  
Qy 1283 TCTTGGAGACTGTGTATCCAGAACTGCATGTGCTGTAGGACAGGCTTGCTTTCGGGA 1342  
Db 62 TCTTGGAGACTGTGTATCCAGAACTGCATGTGCTGTAGGACAGGCTTGCTTTCGGGA 121  
  
Qy 1343 TGCTCCAGAACAAAGGACTAACAGACTACTGCTTTGACTATACCTCCCGATCAAAACC 1402  
Db 122 TGCTCCAGAACAAAGGACTAACAGACTACTGCTTTGACTATACCTCCCGATCAAAACC 181  
  
Qy 1403 AGATTGTGGGACACCAAGTCAATCTGTACTCTGTCTATGGATGGGCGAGAAATCAGTTTT 1462  
Db 182 AGATTGTGGGACACCAAGTCAATCTGTACTCTGTCTATGGATGGGCGAGAAATCAGTTTT 241  
  
Qy 1463 TCGAGTACAGTCCCGAAGAAATACGCTATACACCCACCGCTTGAGGGCTGCATTG 1522  
Db 242 TCGAGTACAGTCCCGAAGAAATACGCTATACACCCACCGCTTGAGGGCTGCATTG 301  
  
Qy 1523 CTGTGGAGCAGGAAATGGATACCTTATCATGCTCTCTCGGAAGAAATGCCCCAGAGA 1582  
Db 302 CTGTGGAGCAGGAAATGGATACCTTATCATGCTCTCTCGGAAGAAATGCCCCAGAGA 361  
  
Qy 1583 ATCAGAAATTCATCTTCAGAGGATGATCTTTATTTTACGAACAGTCCCAAGAAATGTG 1642  
Db 362 ATCAGAAATTCATCTTCAGAGGATGATCTTTATTTTACGAACAGTCCCAAGAAATGTG 421  
  
Qy 1643 TCCAGGCTGCGAGGAAGAGTCCAGTGCAGATTCGTTTCCACTCTTTACGAGACTGCACCA 1702  
Db 1702 TCCAGGCTGCGAGGAAGAGTCCAGTGCAGATTCGTTTCCACTCTTTACGAGACTGCACCA 1702

Db 422 TCCAGCTGCCAGGAGCGAGTGCAGTGCACGTTTCGTTCCACTCTTAGCAGCTGCACCA 481  
Qy 1703 ACTCGATCATCAGAAATGGTCTTCAAGAGCGCATGTTATGA 1746  
Db 482 ACTCGATCATCAGAAATGGTCTTCAAGAGCGCATGTTATGA 525

## RESULT 10

US-09-867-701-4790  
; Sequence 4790, Application US/09867701  
; Patent No. US20020132237A1  
; GENERAL INFORMATION:  
; APPLICANT: Agiate, Paul A.  
; APPLICANT: Jones, Robert  
; APPLICANT: Harlocker, Susan L.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; FILE REFERENCE: 210121.497  
; CURRENT APPLICATION NUMBER: US/09/867,701  
; CURRENT FILING DATE: 2001-05-29  
; NUMBER OF SEQ ID NOS: 10912  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4790  
; LENGTH: 496  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-09-867-701-4790

Query Match 25.8%; Score 450.8; DB 10; Length 496;  
Best Local Similarity 96.8%; Pred. No. 2.7e-116;  
Matches 482; Conservative 0; Mismatches 12; Indels 4; Gaps 2;  
Qy 203 CGCGGTGCCGCGAAGCGCTGGCGCGCGGCGGCGGCGGTGCGGTGCGAGTGCAGG 262  
Db 2 CGCGCGGTGCCGCGAAGCGCTGGCGG--GGAGCGAGGCGGTGCGGTGCGAGTGCAGG 58  
Qy 263 GCGAGGAGCTCGCGCTGCAGAG--GAGAGCTGCGGTGCGACCGATTAACTACCTC 321  
Db 59 GCGAGGAGCTCGCGCTGCAGAGAGAGAGCGTGGACTGCACCGATTAACTACCTC 118  
Qy 322 AGCGACCGCATCTCACTGCACCGCGCTGCCGCGCGCTGCAACCGCTGTGCAAGAG 381  
Db 119 AGCGACCGCATCTCACTGCACCGCGCTGCCGCGCGCTGCAACCGCTGTGCAAGAG 178  
Qy 382 AAGAAATATGATTATGATAATTGCCAGGACATCTGTTATCATAGCATTTTATAATGAA 441  
Db 179 AAGAAATATGATTATGATAATTGCCAGGACATCTGTTATCATAGCATTTTATAATGAA 238  
Qy 442 GCCTGGTCAACTCTCTTCCGACAGTTTACAGTGTCTTGAGACATCCCGGATACCTG 501  
Db 239 GCCTGGTCAACTCTCTTCCGACAGTTTACAGTGTCTTGAGACATCCCGGATACCTG 298  
Qy 502 CTAGAAGAAGTGATCTTGTAGATGACTACAGTGTATAGAGACCTGAGAGGCGCTTG 561  
Db 299 CTAGAAGAAGTGATCTTGTAGATGACTACAGTGTATAGAGACCTGAGAGGCGCTTG 358  
Qy 562 GCCAATGAGCTTTCGGAGCTGCCCAAGTGCGCTGTATCCCGCGCAACAAGAGAGGGC 621  
Db 359 GCCAATGAGCTTTCGGAGCTGCCCAAGTGCGCTGTATCCCGCGCAACAAGAGAGGGC 418  
Qy 622 CTGTGCGAGCCCGGTGCTGGGGCGTCTCGCGGAGGGCGATGTTCTGACCTTCTG 681  
Db 419 CTGTGCGAGCCCGGTCTTCTGGGGCGTCTCGCGGAGGGCGATGTTCTGACCTTCTG 478  
Qy 682 GACTGTCACTGTAGTGC 699  
Db 479 GACTGTCACTGTAGTGC 496

## RESULT 11

US-09-867-701-4789/c  
; Sequence 4789, Application US/09867701  
; Patent No. US20020132237A1

; GENERAL INFORMATION:  
; APPLICANT: Agiate, Paul A.  
; APPLICANT: Jones, Robert  
; APPLICANT: Harlocker, Susan L.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; FILE REFERENCE: 210121.497  
; CURRENT APPLICATION NUMBER: US/09/867,701  
; CURRENT FILING DATE: 2001-05-29  
; NUMBER OF SEQ ID NOS: 10912  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4789  
; LENGTH: 473  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-09-867-701-4789

Query Match 25.4%; Score 443.4; DB 10; Length 473;  
Best Local Similarity 99.4%; Pred. No. 3.1e-114;  
Matches 466; Conservative 0; Mismatches 1; Indels 2; Gaps 2;  
Qy 595 CTGATCCGCGCCAAAGAGAGAGGGCTGTCGAGCCCGGCTGCTGGGGGCTCTGCG 654  
Db 473 CTTATCCGCGCCAAAGAGAGAGGGCTGTCGAGCCCGGCTGCTGGGGGCTCTGCG 414  
Qy 655 GCGAGGGCGGATGTTCTGACCTTCTGACTGTCTGAGTGCCTGAGTGCAGAGGGTGGCTG 714  
Db 413 GCGAGGGCGGATGTTCTGACCTTCTGACTGTCTGAGTGCCTGAGTGCAGAGGGT-GCTG 355  
Qy 715 GAGCCGCTGCTGCAGAGGATCCATGAAGAGAGTCCGAGTGGTGTGCTGCCGCTGATTGAT 774  
Db 354 GAGCCGCTGCTGCAGAGGATCCATGAAGAGAGTCCGAGTGGTGTGCTGCCGCTGATTGAT 295  
Qy 775 GTGATCGACTGGAACACCTTCGAAATACCTCGGGAACCTCCGGGAGCCCGAGATCGGCGGT 834  
Db 294 GTGATCGACTGGAACACCTTCGAAATACCTCGGGAACCTCCGGGAGCCCGAGATCGGCGGT 235  
Qy 835 TTCGACTGGAGGCTGGTGTTCAGC-TGGCACACAGTTTCTTGAGAGGAGAGGATACCGAT 893  
Db 234 TTCGACTGGAGGCTGGTGTTCAGC-TGGCACACACAGTTTCTTGAGAGGAGAGGATACCGAT 175  
Qy 894 GCATCCCGCTGCAGTGTCTCAGGTCTCCAAATGGCTGGTGGGCTGTTTGTGTGAG 953  
Db 174 GCATCCCGCTGCAGTGTCTCAGGTCTCCAAATGGCTGGTGGGCTGTTTGTGTGAG 115  
Qy 954 TAAGAATAATTTGAATATCTGGGTCTTATGATACAGGAATGGAAGTTTGGGAGGAGA 1013  
Db 114 TAAGAATAATTTGAATATCTGGGTCTTATGATACAGGAATGGAAGTTTGGGAGGAGA 55  
Qy 1014 AAACCTCGAATTTTCTTTAGGATCTGGCAGTGTGGTGGGTTCTCGAA 1062  
Db 54 AAACCTCGAATTTTCTTTAGGATCTGGCAGTGTGGTGGGTTCTCGAA 6

## RESULT 12

US-09-867-701-3683/c  
; Sequence 3683, Application US/09867701  
; Patent No. US20020132237A1  
; GENERAL INFORMATION:  
; APPLICANT: Agiate, Paul A.  
; APPLICANT: Jones, Robert  
; APPLICANT: Harlocker, Susan L.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; FILE REFERENCE: 210121.497  
; CURRENT APPLICATION NUMBER: US/09/867,701  
; CURRENT FILING DATE: 2001-05-29  
; NUMBER OF SEQ ID NOS: 10912  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3683  
; LENGTH: 473  
; TYPE: DNA  
; ORGANISM: Homo sapien

US-09-867-701-3683

```
Query Match          24.4%; Score 425.8; DB 10; Length 473;
Best Local Similarity 98.7%; Pred. No. 2.8e-109;
Matches 471; Conservative 0; Mismatches 2; Indels 4; Gaps 4;

Qy 592 CGCTGATCCGCGCAACAAGAGAGAGGGCCGTGGAGCCGCGCTGCTGGGGCGTCT 651
Dy 473 CGCTGATCCGCGCAACAAGAGAGAGGGCCGTGGAGCCGCGCTGCTGGGGCGTCT 414

Qy 652 CGCGGAGGGCGGATGTTCTGACCTTCTGACCTGTGACCTGTGACCTGAGTGCACGAAGGTGG 711
Dy 413 CGCGCGA-GGGCGATGTTCTGACCTTCTGACCTTCTGACCTGTGACCTGAGTGCACGAAGGT-G 356

Qy 712 CTGAGCGCGCTGTCGAGAGGATCCATGAAGAGGTCGGCAGGTGTCGCCGGTGATT 771
Dy 355 CTGAGCGCGCTGTCGAGAGGATCCATGAAGAGGTCGGCAGGTGTCGCCGGTGATT 296

Qy 772 GATGTGATCGACTGGAAACACCTTCGAATACCTCGGGGAACCTCGGGGAGCCCGACATCGGC 831
Dy 295 GATGTGATCGACTGGAAACACCTTCGAATACCTCGGGGAACCTCGGGGAGCCCGACATCGGC 237

Qy 832 GGTTCGATCGAGGCTGGTGTTCAGTGGCACACAGTTCCTGAGAGGAGGATACGG 891
Dy 236 GGTTCGATCGAGGCTGGTGTTCAGTGGCACACAGTTCCTGAGAGGAGGATACGG 177

Qy 892 ATGCAATCCCGCTGATGTCATCAGGTCTCCAAATGCTGCTGGGCTGTTGCTGTG 951
Dy 176 ATGCAATCCCGCTGATGTCATCAGGTCTCCAAATGCTGCTGGGCTGTTGCTGTG 117

Qy 952 AGTAAGAAATATTTGAATATCTCGGGTCTTATGATACAGGAATGGAAGTTTGGGGAGGA 1011
Dy 116 AGTAAGAAATATTTGAATATCT-GGGTCTTATGATACAGGAATGGAAGTTTGGGGAGGA 58

Qy 1012 GAAACCTCGAATTTCTTTAGATCTGCGAGTCTGCGAGTGTGGTGGGTTCTGGAACAC 1068
Dy 57 GAAACCTCGAATTTCTTTAGATCTGCGAGTGTGGTGGGTTCTGGAACAC 1

RESULT 13
US-09-777-564-414
; Sequence 414, Application US/09777564
; Patent No. US20020022591A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.493
; CURRENT APPLICATION NUMBER: US/09/777,564
; CURRENT FILING DATE: 2001-02-05
; NUMBER OF SEQ ID NOS: 1730
; SOFTWARE: FastSeq for Window Version 4.0
; SEQ ID NO 414
; LENGTH: 544
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(544)
; OTHER INFORMATION: n = A,T,C or G
US-09-777-564-414
```

```
Query Match          24.0%; Score 418.4; DB 9; Length 544;
Best Local Similarity 99.8%; Pred. No. 3.5e-107;
Matches 419; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1327 CCTGGCTTCTTCGGGATGCTCCAGAACAAAGACTTAACAGACTACTGCTTTGACTATAAC 1386
Dy 1 CCTGGCTTCTTCGGGATGCTCCAGAACAAAGACTTAACAGACTACTGCTTTGACTATAAC 60

Qy 1387 CCTCCCGATGAAACACAGATTGTGGACACACAGGTCACTTGTACCTCTGTCATGGGATG 1446
```

```
Db 61 CCTCCCGATGAAACACAGATTGTGGGACACCAAGTCATTCTGTACCTCTCTCATGGGATG 120
Qy 1447 GGCAGAAATCAGTTTTTTCGAGTACACGTCCAGAGAAATAACGCTATATAACACCCACAG 1506
Dy 121 GGCAGAAATCAGTTTTTTCGAGTACACGTCCAGAGAAATAACGCTATATAACACCCACAG 180
Qy 1507 CTTGAGGGCTGCAATGCTGTGGAGCAGGAATCGATACCCCTTATCATGCACTCTCTGGAA 1566
Dy 181 CTTGAGGGCTGCAATGCTGTGGAGCAGGAATCGATACCCCTTATCATGCACTCTCTGGAA 240
Qy 1567 GAAACTGCCCCCAGAGAAATCAGAAGTTTCATCTTCAGAGAGATGATCTTTATTTACGAA 1626
Dy 241 GAAACTGCCCCCAGAGAAATCAGAAGTTTCATCTTCAGAGAGATGATCTTTATTTACGAA 300
Qy 1627 CAGTCCAAGAAATGTGTCCAGGCTGCGAGGAAGAGTCGAGTGCACAGTTTCGTTCCACTC 1686
Dy 301 CAGTCCAAGAAATGTGTCCAGGCTGCGAGGAAGAGTCGAGTGCACAGTTTCGTTCCACTC 360
Qy 1687 TTACGAGACTGCACCAACTCGGATCATCAGAAATGGTTCTTCAAAGAGCGCATGTTATGA 1746
Dy 361 TTACGAGACTGCACCAACTCGGATCATCAGAAATGGTTCTTCAAAGAGCGCATGTTATGA 420

RESULT 14
US-10-015-219-414
; Sequence 414, Application US/10015219
; Publication No. US20030008299A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.493C1
; CURRENT APPLICATION NUMBER: US/10/015,219
; CURRENT FILING DATE: 2002-03-02
; NUMBER OF SEQ ID NOS: 1739
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 414
; LENGTH: 544
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 544
; OTHER INFORMATION: n = A,T,C or G
US-10-015-219-414
```

```
Query Match          24.0%; Score 418.4; DB 14; Length 544;
Best Local Similarity 99.8%; Pred. No. 3.5e-107;
Matches 419; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1327 CCTGGCTTCTTCGGGATGCTCCAGAACAAAGACTTAACAGACTACTGCTTTGACTATAAC 1386
Dy 1 CCTGGCTTCTTCGGGATGCTCCAGAACAAAGACTTAACAGACTACTGCTTTGACTATAAC 60

Qy 1387 CCTCCCGATGAAACACAGATTGTGGGACACACAGGTCACTTGTACCTCTGTCATGGGATG 1446
Dy 61 CCTCCCGATGAAACACAGATTGTGGGACACACAGGTCACTTGTACCTCTGTCATGGGATG 120
Qy 1447 GGCAGAAATCAGTTTTTTCGAGTACACGTCCAGAGAAATAACGCTATAACACCCACAG 1506
Dy 121 GGCAGAAATCAGTTTTTTCGAGTACACGTCCAGAGAAATAACGCTATAACACCCACAG 180
Qy 1507 CTTGAGGGCTGCAATGCTGTGGAGCAGGAATGATACCTTATCATGCACTCTCTGGAA 1566
Dy 181 CTTGAGGGCTGCAATGCTGTGGAGCAGGAATGATACCTTATCATGCACTCTCTGGAA 240
Qy 1567 GAAACTGCCCCCAGAGAAATCAGAAGTTTCATCTTCAGAGAGATGGAATCTTTATTTACGAA 1626
Dy 241 GAAACTGCCCCCAGAGAAATCAGAAGTTTCATCTTCAGAGAGATGGAATCTTTATTTACGAA 300
Qy 1627 CAGTCCAAGAAATGTGTCCAGGCTGCGAGGAAGAGTCGAGTGCACAGTTTCGTTCCACTC 1686
Dy 301 CAGTCCAAGAAATGTGTCCAGGCTGCGAGGAAGAGTCGAGTGCACAGTTTCGTTCCACTC 360
```

Qy 1687 TTACGAGACTGCACCACTCGGATCATCAGAAATGGTTCTTCAAAGAGCGCATGTTATGA 1746  
Db 361 TTACGAGACTGCACCACTCGGATCATCAGAAATGGTTCTTCAAAGAGCGCATGTTATGA 420

## RESULT 15

US-09-777-564-172  
; Sequence 172, Application US/09777564  
; Patent No. US20020022591A1  
; GENERAL INFORMATION:  
; APPLICANT: Mannion, Paul A.  
; APPLICANT: Algate, Paul A.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER  
; FILE REFERENCE: 210121.493  
; CURRENT APPLICATION NUMBER: US/09/777,564  
; CURRENT FILING DATE: 2001-02-05  
; NUMBER OF SEQ ID NOS: 1730  
; SOFTWARE: FastSeq for Window Version 4.0  
; SEQ ID NO 172  
; LENGTH: 632  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(632)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-777-564-172

Query Match 22.9%; Score 400.2; DB 9; Length 632;  
Best Local Similarity 98.1%; Pred. No. 4.9e-102;  
Matches 413; Conservative 0; Mismatches 7; Indels 1; Gaps 1;  
Qy 1327 CCTGGCTTCTTCGGGATGCTCCAGACAAAGGACTTAACAGACTACTGCTTTTGACTATAAC 1386  
Db 1 CCTGGCTTCTTCGGGATGCTCCAGACAAAGGACTTAACAGACTACTGCTTTTGACTATAAC 60  
Qy 1387 CCTCCCGATGAAACCAAGATTGTGGGACACAGGTCTATCTGTATCTCTGTATGGGATG 1446  
Db 61 CCTCCCGATGAAACCAAGATTGTGGGACACAGGNCATCTGTACCTCTGTATGGGATG 120  
Qy 1447 GGCCAGATC-AGTTTTCGAGTACAGTCCAGAAAGAAATACGCTATAACCCACCA 1505  
Db 121 GGCCAGATCAAGTTTTCGAGTACACTTCCAGAAAGAAATACGCTATAACCCACCA 180  
Qy 1506 GCCTGAGGGCTGCTTGTGGAGCAGGAATGATACCTTATCATGCTCTCTCGGA 1565  
Db 181 GCCTGAGGGCTGCTTGTGGAGCAGGAATGATACCTTATCATGCTCTCTCGGA 240  
Qy 1566 AGAACTGCCCCAGAGAAATCAGAAAGTTTCATCTTCAGGAGGATGGATCTTTATTTACGA 1625  
Db 241 AGAACTGCCCCAGAGAAATCAGAAAGTTTCATCTTCAGGAGGATGGATCTTTATTTACGA 300  
Qy 1626 ACAGTCAAGAAATGCTCCAGGCTGGAGGAGGATCGAGTGACAGTTTCGTTCCACT 1685  
Db 301 ACAGTCAAGAAATGCTCCAGGCTGGAGGAGGATCGAGTGACAGTTTCGTTCCACT 360  
Qy 1686 CTTACGAGACTGCACCAACTCGGATCATCAGAAATGGTTCTTCAAAGAGCGCATGTTATG 1745  
Db 361 CTTACGAGACTGCACCAACTCGGATCATCAGAAATGGTTCTTCAAAGAGCGCATGTTATG 420  
Qy 1746 A 1746  
Db 421 A 421

Search completed: December 8, 2003, 01:56:05  
Job time : 534.474 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 7, 2003, 12:46:13 ; Search time 3629.14 Seconds  
(without alignments)  
11693.015 Million cell updates/sec

Title: US-10-074-527-3

Perfect score: 1746

Sequence: 1 atgtggggggcgcgcgcg.....tcaaagagcgcatttatga 1746

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

BST.\*

1: em\_estba.\*

2: em\_estum.\*

3: em\_estin.\*

4: em\_estmu.\*

5: em\_estov.\*

6: em\_estpl.\*

7: em\_estro.\*

8: em\_htc.\*

9: gb\_est1.\*

10: gb\_est2.\*

11: gb\_htc.\*

12: gb\_est3.\*

13: gb\_est4.\*

14: gb\_est5.\*

15: em\_estfun.\*

16: em\_estom.\*

17: em\_gss\_hum.\*

18: em\_gss\_inv.\*

19: em\_gss\_pln.\*

20: em\_gss\_vrt.\*

21: em\_gss\_fun.\*

22: em\_gss\_mam.\*

23: em\_gss\_mus.\*

24: em\_gss\_pro.\*

25: em\_gss\_rod.\*

26: em\_gss\_phg.\*

27: em\_gss\_vrl.\*

28: gb\_gss1.\*

29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1249.4	71.6	2074	11 AK042133	AK042133 Mus muscu
2	1249.4	71.6	2217	11 AK033638	AK033638 Mus muscu
3	1238.4	70.9	2034	11 AK082014	AK082014 Mus muscu
4	1025.4	58.7	1781	11 BC025639	BC025639 Mus muscu

# ALIGNMENTS

RESULT 1

AK042133

LOCUS

DEFINITION

AK042133

LOCUS

DEFINITION

AK042133

LOCUS

DEFINITION

AK042133

LOCUS

DEFINITION

AK042133

LOCUS

DEFINITION

AK042133

LOCUS

DEFINITION

AK042133

LOCUS

DEFINITION

AK042133

LOCUS

DEFINITION

AK042133

LOCUS

DEFINITION

AK042133

LOCUS

DEFINITION

5 936.2 53.6 2064 11 BC024475  
6 925.2 53.0 1752 11 AK043977  
7 925.2 37.8 676 10 BE677813  
8 566.4 32.4 692 14 CB165548  
9 568.2 31.4 561 9 AI800923  
10 520.8 29.8 2607 11 AK033494  
11 519 29.7 967 10 BG167520  
12 508 29.1 661 12 BI154821  
13 503.2 28.8 721 12 BG868185  
14 480.6 27.5 511 9 AI863865  
15 450.8 25.8 496 9 AA429394  
16 448.2 25.7 700 13 BU307661  
17 443.4 25.4 473 9 AA429393  
18 426.2 24.4 530 10 BG608482  
19 425.8 24.4 473 9 AA401053  
20 424.8 24.3 653 10 BG625226  
21 424.2 24.3 457 12 BG993195  
22 413 23.7 664 10 BG638950  
23 387.4 22.2 477 10 BG608487  
24 385.8 22.1 628 10 BG649873  
25 373.4 21.4 521 10 BE647828  
26 373 21.4 544 4 BX514155  
27 365.4 20.9 697 10 BB272149  
28 352.8 20.2 772 10 BF247687  
29 347.4 19.9 703 12 BI663830  
30 346.8 19.9 462 14 CD554038  
31 340 19.5 445 10 BF211275  
32 335 19.2 407 12 BI403200  
33 333.8 19.1 909 14 CD251293  
34 332.2 19.0 441 9 AW462202  
35 331.8 19.0 464 13 BQ829759  
36 329.6 18.9 458 10 BE138172  
37 325 18.6 445 14 CB552134  
38 323 18.5 690 10 BF031122  
39 322.2 18.5 360 9 AA493187  
40 320.4 18.4 442 12 BM254405  
41 318.4 18.2 818 10 BF246502  
42 311.8 17.9 634 12 BJ067676  
43 306.4 17.5 471 9 AA871181  
44 298 17.1 947 10 BG294191  
45 296 17.0 770 14 CD299534

AK042133 Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched library, clone:A630062B03 product:hypothetical Glycosyl transferase, family 2 containing protein, full insert sequence.

AK042133 GI:26334956

HTC; CAP trapper.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 Carninci, P., Shibata, Y., Hayashizaki, Y., Sugahara, Y., Shibata, K.,

Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning

Meth. Enzymol. 303, 19-44 (1999)

9279253

10349636

2

Carninci, P., Shibata, Y., Hayashizaki, Y., Sugahara, Y., Shibata, K.,

Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes

Genome Res. 10 (10), 1617-1630 (2000)

20499374

11042159



Db 519 GAGACCTCCCTGACATCTCTGCTGAGAGAGTCAATCTTGTTAGATGACTACAGCGACAGA 578  
Qy 541 GAGCACCTGAGGAGCGCTTGGCCAAATGAGCTTTTCGGGACTGCGCCAAAGGTGCGCTGATC 600  
Db 579 GAGCACCTAAAGGAACGCTTGGCCAAACGAGCTGTACAGCTCCCAAGGTGCGCTGATC 638  
Qy 601 CGCGCCAAAGAGAGAGGCGCTGCTGCGAGCCCGGCTGTGCGGGGCGTCTGCGGCGAGG 660  
Db 639 CGTGCTAGCAGGAGAGAGGCGCTAGTGCAGCCCGGCTACTGCGGAGCGCTCTGCGGCGCAGG 698  
Qy 661 GCGGATGCTTCGACCTTCCTGAGCTGTCACTGTGAGTGCACGAAGGGTGGCTGGAGCGG 720  
Db 699 GCGGAGTGTGACCTTCTTAGACTGTCACTGTGAATGTATGAGGGGTGGCTGGAGCCT 758  
Qy 721 CTGCTGAGAGGATCCATGAAGAGGAGTCCGCGAGTGTGCTGCCGCTGATTGTGATC 780  
Db 759 CTGCTGAGAGGATCCACGAGAGGAGTCCGCGAGTGTGCTGCCCTGTTATCGATGTGAT 818  
Qy 781 GACTGGAACACCTTCGAAATACCTGGGGAATCTCGGGGAGCCCGAGATCGGGGTTTCGAC 840  
Db 819 GACTGGAACACCTTTGAGTACTTGGGCAACTCGGGAGAGCGCGAGATTGGCGGCTTTGAC 878  
Qy 841 TGAGGCTGTGTTTCACTGTCGCACACAGTTCTCGAGAGGAGAGGATACGGATGCAATCC 900  
Db 879 TGGCGGCTGTATTCAGTGGCAGCTGTGCTCCCGAGCGGAGCGGAGTGCATGCGGTG 938  
Qy 901 CCGTCTGATGTCATCAGTCTCCAAACATGGCTGGTGGGCTGTTGCTGTGAGTAAAGAA 960  
Db 939 CCGATCGATGTATCAGTCTCCAACTATGCTGGGGGCTGTTGCTGTGAGTAAAGAA 998  
Qy 961 TATTTTGAATATCTGGGCTTATGATACAGGAATGAAGTTGGGGAGGAGAAACCTC 1020  
Db 999 TATTTTGAATATCTGGGCTTATGATACAGGAATGAAGTTGGGGAGGAGAAACCTT 1058  
Qy 1021 GAATTTTCTTTAGGATCTGCGAGTGTGGGTTCTGGAAACACACCCATGTTCCCAT 1080  
Db 1059 GAGTTCCTTTAGGATCTGCGAGTGTGGTGGCACTCTGGAACGCGACCCGTGCTCCAT 1118  
Qy 1081 GTTGGCCATGTTTCCCAAGCAAGCTCCCTACTCCCGCAACAGGCTCTGGCCAAACAGT 1140  
Db 1119 GTGGCCACGCTTCCCTAGCAAGCTCCCTATTCCTGAGCAAGGCTTGGCCAAACAGT 1178  
Qy 1141 GTTCTGTCAGCTGAAGTATGATGATGAATTTAAAGAGCTCTACTACCATCGCAACCCC 1200  
Db 1179 GTCCGAGCTGAGAGGTGTGATGATGAATTTAAAGAACTCTACTACCAACGCAATCCC 1238  
Qy 1201 CGTCCCGCTTGAACCTTTTGGGATGTACAGAGAGGAGCGAGCTCCGGGCAAGCTC 1260  
Db 1239 CAGGCCCGCTTGAACCTTTTGGGAGCTGACAGAGAGGAGAGCTTCGGGCTTAAGCTC 1298  
Qy 1261 CAGTGTAAAGACTTCAAGTGGTCTTGGAGACTGTGTATCCAGAACTGCATGTGCTTGAG 1320  
Db 1299 CAGTGTAAAGACTTCAAGTGGTCTTGAATACAGTGTACCAAGATTGCAGTGCACAGAG 1358  
Qy 1321 GACAGGCTGGCTTCTTCGGAGTCTCCAGAAACAAAGGACTTAACAGACTACTGCTTTGAC 1380  
Db 1359 GACAGGCTGGCTTCTTCGGAGTCTCCAGAAACAGAGGCTTAAGAGGCTACTGCTTTGAC 1418  
Qy 1381 TATAACCTCCGATGAAACACAGATTGTGGGACACAGGTCATTCTGTACCTCTGTGAT 1440  
Db 1419 TACAATCTCCCAATGAAACCAATGTTGAAGGGCCACAGGTCCTTCTGTACCTCTGCGAT 1478  
Qy 1441 GCGATGGCCAGAAATCAGTTTTTTCGAGTACAGTCCAGAAAGAAATACGCTATAACACC 1500  
Db 1479 GCGATGGCTCAGAACCAAGTTTTTCGAGTATACGACCCGGAAGAAATACGCTATAACACC 1538  
Qy 1501 CACCAGCTGAGGCTGCATTTGCTGTGGAAGCAGGAATGGATACCTTTATCATGATCTC 1560  
Db 1539 CGCCAGCCCGAAGCTGCATCACCGTGGAGGACGGGAGGATACCTTTGTGATGGATCTC 1598  
Qy 1561 TGGAGAAATCTGCCCCAGAGAAATCAGAAATTCAGATTTCTTTCAGGAGGATGATCTTTATTT 1620

Db 1599 TGCGGAGAGACCGTCCAGAGAACACAGGAGTTTCATCTCAGAGGAGATGCGACTTTAGTT 1658  
Qy 1621 CACGAACAGTCCAAAGAAATGTGTCCAGGCTCGAGGAAGAGTGCAGTGACAGTTTCGTT 1680  
Db 1659 CACAAGCACAGCAGGAATGTGTGGAGGCCACAGAGAGGTGTAGACAACGCGTTTGCA 1718  
Qy 1681 CCACTCTTAGGAGACTGCACCACTCGGATCATCAGAAATGGTTCTTCAAAGAGCGCATG 1740  
Db 1719 CCATCTTAGCGGACTGTACCACTCAGATAACAGAGGTGGTTCTTCAAAGGAGCGCATG 1778  
Qy 1741 T 1741  
Db 1779 T 1779

AK033638 2217 bp mRNA linear HTC 05-DEC-2002  
Mus musculus adult male cecum cDNA, RIKEN full-length enriched  
library, clone:1910206E10 product:hypothetical Glycosyl  
transferase, family 2 containing protein, full insert sequence.  
AK033638  
ACCESSION AK033638.1 GI:26329324  
VERSION  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
REFERENCE 1  
AUTHORS Carninci, P. and Hayashizaki, Y.  
TITLE High-efficiency full-length cDNA cloning  
JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
MEDLINE 99279253  
PUBMED 10349636  
REFERENCE 2  
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
MEDLINE 20493374  
PUBMED 11042159  
REFERENCE 3  
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Konno, H., Akiyama, J., Nishi, K., Kiteunai, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
Yamanoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
TITLE RIKEN integrated sequence analysis (RISA) system -384-format  
sequencing pipeline with 384 multicapillary sequencer  
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)  
MEDLINE 20530913  
PUBMED 11076861  
REFERENCE 4  
AUTHORS Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,  
Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,  
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I.,  
Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,  
Kadota, K., Matsuoka, H., Ashburner, M., Batalov, S., Casavant, T.,  
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,  
Kuehli, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G.,  
Quackenbush, J., Schriml, L.M., Staehli, F., Suzuki, R., Tomita, M.,  
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,  
Baldarelli, R., Barsh, G., Blake, J., Boiffelli, D., Bojunga, N.,  
Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C.,  
Fleischer, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,  
Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P.,  
Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P.,  
Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,  
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H.,  
Toyooka, K., Wang, K.H., Weitz, C., Whittaker, C., Wilming, L.,

Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S., and Hayashizaki, Y.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409 (5821), 685-690 (2001)  
21085660  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
5  
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
6 (bases 1 to 2217)  
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hangaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohashi, N., Okazaki, Y., Saio, R., Saichou, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.  
Direct Submission  
Submitted (15-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan [E-mail: genome-res@gsr.riken.go.jp].  
URL: <http://genome.gsc.riken.go.jp/>, Tel: 81-45-503-9222, Fax: 81-45-503-9216  
CDS library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Please visit our web site for further details.  
URL: <http://genome.gsc.riken.go.jp/>  
URL: <http://fantom.gsc.riken.go.jp/>.  
Location/Qualifiers  
1. :2217  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM\_DB:9130206E10"  
/db\_xref="taxon:10090"  
/clone="9130206E10"  
/sex="male"  
/tissue type="cecum"  
/clone\_libs="RIKEN full-length enriched mouse cDNA library"  
/dev stages="adult"  
22. :1752  
/note="unnamed protein product; hypothetical Glycosyl transferase, family 2 containing protein (InterPro|IPR001173, evidence: InterPro)  
putative"  
/codon\_start=1  
/protein\_id="BAC28401.1"  
/db\_xref="GI:26329325"  
/translation="MWGRAVRRRCPRGLRGREALLALLAGLALLRARSRSRTVD  
EGPRTPLGRHEPVPPLPADALGNHGEAVPLQGBELRLQESVVKQHNIYL  
SDRLSLRLHRLERNWPLCKEVKDYDNLPTKSVVIAFNEMWSTLLRTVYSVLSTPD  
ILLBEVLDDVDSYDREHLKERLANELSQLPKVIRASRRGLVRALLGASAAEVS  
LFLDCEHCEHGLEFLQRIHEKSAVVCPIVDVDMNTFEYLGNSGPEQIGGFOW  
RLVFTWVPPORQSRSPIDVIRFPTMAGGLFAVSKRVPDYLGSDYDNGVNGEN  
LEFSRTWOCQGTLETSPCHSVGHVFPKQAPYSRSKALANSVRAAEVMDFEKELYH  
RNPOARLEPFGDVTKRLAKLQCKDFKFLDTPVPELVHPEDRPFGLMQRGLR  
GYLDYFNPNENHVEGHVQLLYLCHMGQNGFFETTKRKEIRYNTRPFACTIVEDGR  
DTLYMDLCRFTVPENQBFILQEDGTLVHKHSRKCKVEATEKVLNDNGFAPYLRDCTNSDN  
QRWFFKERM"  
2198. :2203  
polya\_signal

polya\_site /note="putative"  
2217  
/note="putative"  
BASE COUNT 503 a 543 c 666 g 505 t  
ORIGIN  
Query Match 71.6%; Score 1249.4; DB 11; Length 2217;  
Best Local Similarity 83.3%; Pred. No. 8.1e-270;  
Matches 1450; Conservative 0; Mismatches 276; Indels 15; Gaps 2;  
Qy 1 ATGTGGGGGCGCAGCGCGCGCGCTGCGCGGGAACTGCGCGCGCGCGCGGAGCG 60  
Db 22 ATGTGGGGGCGCTGGGTGAGAGGCGCTGCGCGGGGGTGGCGCGCGCGGAGCG 81  
Qy 61 CTGTTGGTGTCTCTGGCGCTACTGGCGTTGGCGGGGCTGGGTGCTGGCGGCGCAG 120  
Db 82 TTG-----CTGGCACTGCTGGCGCTGGCGGGCTGGCGGGCTGCTGGCGGC--- 128  
Qy 121 CGTGGGGCGCGGGGCTGGCGAGCGCGGACCCCGCGGCACCCCGCGCGCGCGCGCGG 180  
Db 129 --TCGAAGCAGGTGGGGACGGTGGACCCCGGGACCCCGCGGACTCTCTCCCGGGCGT 186  
Qy 181 CGCGAGCGGTCTATGCGCGCGCGCGGTGCGCGGCGAAACGCGCTGGCGCGCGCGGCGAG 240  
Db 187 CACGAGCAGTGTGCGCGCGCGCGCTGCGCGCGCGCTCTGGGAGGCGCATGGGGAG 246  
Qy 241 GCGGTGGCGCTGAGCTGCGAGGCGGAGGCTGCGGCTGCGAGGAGGAGCGTGGCGCTG 300  
Db 247 GCGGTGGCGCTGAGCTGCGAGGCGGAGGCTGCGGCTGCGAGGAGGAGCGTGAAGCAG 306  
Qy 301 CACGAGATTAACTACTACCTCAGCGACCGCATCTCACTGACCGCGCGCGCTGCCCGAGCGC 360  
Db 307 CACGAGATCAACATCTATCTGAGGACCGTATCTGCTGCGCGCGGACTGCCCGAGCGC 366  
Qy 361 TGAACCGCGCTGTGCAAGAGAGAAATATGATATGATTAATTTGCCAGAGACATCTGTT 420  
Db 367 TGAACCGCTGTGCGAGAGAGGTGAATACGATTATGATAACCTGCCCAAGACTTCTGTT 426  
Qy 421 ATCATAGCATTTTATATGAAGCGCTGGTCACTCTCTCGGACAGTTTACAGTGTCTT 480  
Db 427 GTCATAGCATTTTATATGAAGCGCTGGTCCACACTCTCTCGGACAGTTTACAGTGTCTT 486  
Qy 481 GAGACATCCCCGGATATCTGCTAGAGAAAGTATCTTGTAGATGACTACAGTGATAGA 540  
Db 487 GAGACCTCCCTGCATCTCTGCTGGAGAGGTCACTTCTGTAGATGACTACAGCGACAGA 546  
Qy 541 GAGCCTCTGAAGAGAGCGCTTTGGCAATGAGCTTTGGGAGCTGCCCAAGGTGGCGCTGATC 600  
Db 547 GAGCACCCTAAAGGAAACGCTTTGGCCACACGAGCTGTACAGCTCCCAAGGTGGCGCTGATC 606  
Qy 601 CGCGCCCAAGAGAGAGGCGCTGTCGAGCGCGCGCTGTCGGGGCGTCTGGCGCGAGG 660  
Db 607 CGTGTAGCAGAGAGAGGCGCTAGTGTGAGCGCGCGCTACTGGGAGCGCTCTGGCGCGAGG 666  
Qy 661 GGCAGTGTCTGACCTTCTCTGAGTGTCACTGTGAGTGCACAGCAAGGCTGGCTGGAGCGG 720  
Db 667 GGCAGTGTCTGACCTTCTCTGAGTGTCACTGTGAGTGTCACTGTGAGGCTGGCTGGAGCGCT 726  
Qy 721 CTGCTGAGAGGATCCATGAAGAGGAGTGGCGAGTGTGTGTCGGGTGATTTGATGTGATC 780  
Db 727 CTGCTGAGAGGATCCACAGAGAGGAGTGGCGAGTGTGTGTCGGGTGATTTGATGTGATTT 786  
Qy 781 GACTGGCAACCTTCGAAATACCTGGGGAACCTCGGGGAGCGCCAGATCGCGGTTTCGAC 840  
Db 787 GACTGGCAACCTTTGAGTACCTGGGCAACTCGGGAGAGCGCGAGATTTGGCGGCTTTGAC 846  
Qy 841 TGGAGGCTGTGTGTTTCACTGGGCAACACTTCTTCTGAGAGGAGGAGATACGAGATCAATCC 900  
Db 847 TGGCGGCTGTGATTTACGTGCGACGTGTTTCCCGAGCGGAGCGGCGAGTCAATCGGCTG 906  
Qy 901 CCGCTCGATGCTATCAGGTCTTCAAACTATGGTGGTGGGTGTTTGTGTGTGAGTAAGAAA 960  
Db 907 CCGATCGATGTTATCAGGTCTTCAAACTATGGTGGGCGCGCTGTTTGTGTGTGAGTAAGAGA 966

Qy	961	TATTTTGAATATCTGGGGTCTTATGATACAGGAATGGAAGTTTGGGAGGAGAAAAACCTC	1020
Ds	967	TATTTTGAATATCTGGGGTCTTATGATACAGGAATGGAAGTTTGGGAGGAGAAAAACCTT	1026
Qy	1021	GAATTTTCTTTAGGATCTGGCAGTGTGGGGTCTTGGAAACACACCCCATGTCCTCAT	1080
Ds	1027	GAGTTCTCTTTAGGATCTGGCAGTGTGGGGTCTTGGAAACACCCCATGTCCTCAT	1086
Qy	1081	GTTGGCCATGTTTCCCAAGCAAGCTCCCTACTCCCGCAACAAGGCTCTGGCCAAACAGT	1140
Ds	1087	GTGGCCACAGTCTTCCCTAAGCAAGCTCCCTACTCCCGCAACAAGGCTCTGGCCAAACAGT	1146
Qy	1141	GTTCTGTGAGCTGAAGTATGATGATGAATTTAAAGAGCTCTACTACCATCGCAACCCC	1200
Ds	1147	GTCCGAGCTGCAGAGTGTGGATGATGAATTTAAAGAGCTCTACTACCATCGCAACCCC	1206
Qy	1201	CGTCCCGCTTGGAACTTTTGGGATGTGACAGAGGAGCAGCTCCGGCAAGCTC	1260
Ds	1207	CAGCCCGCTTGGAACTTTTGGGATGTGACAGAGGAGCAGCTTCGGGCTTAAGCTC	1266
Qy	1261	CAGTGAAGACTTCAAGTGGTCTTGGAGACTGTGTATCCAGAACTGCATGTGCCTGAG	1320
Ds	1267	CAGTGAAGACTTCAAGTGGTCTTGGAGACTGTGTATCCAGAACTGCATGTGCCTGAG	1326
Qy	1321	GACAGGCTGCTTCTTGGGATGCTCCAGAAACAGGACTTAACAGACTACTGCTTTGAC	1380
Ds	1327	GACAGGCTGCTTCTTGGGATGCTCCAGAAACAGGACTTAACAGACTACTGCTTTGAC	1386
Qy	1381	TATAACCTCCCGATGAACACCATGTTGGGACACAGGCTCATCTGTACCTCTGTCAT	1440
Ds	1387	TACAATCTCCCAATGAACACCATGTTGAAGGCCACACAGGCTCTTGTACCTCTGTCAT	1446
Qy	1441	GGATGGGCCAGAACTAGTCTTTCAGTACACGCTCCAGAAAGAAATACGCTATAACACC	1500
Ds	1447	GGATGGGTGACAGACCATGTTTCAGTATACGACCCGGAAGAAATACGCTATAACACC	1506
Qy	1501	CACAGGCTGAGGCTGATGCTGTGGAAGCAGGAATGGAATACCTTATCATCATCTC	1560
Ds	1507	CGCCAGGCGGAAGCTGCATCACCCTGGAGCAGGGAAGGATACCTTGTTCATGGATCTC	1566
Qy	1561	TGCGAAGAACTGCCAGAGAACTCAGAGTTCATCTTCGAGGAGGATGATCTTTATTT	1620
Ds	1567	TGCGAGAGACCGTCCCAAGAACACGAGGTTTCATCTACAGGAGGATGACATTTAGTT	1626
Qy	1621	CACGAACAGTCCAGAAATGTCTCAGGCTCGAGGAGGAGTCCAGTGCAGTTTCGTT	1680
Ds	1627	CACAGCACAGCAAGAAATGTCTGAGGCCACAGAGAGGTGTAGACACGGCTTTGCA	1686
Qy	1681	CCACTCTTACGAGACTGCACCAACTCGGATCATCAGAAATGGTTCTTCAAAGAGCGCATG	1740
Ds	1687	CCATCTTACGGGACTGTACCAACTCAGATAACAGAGGTGGTTCTTCAAAGAGCGCATG	1746
Qy	1741	T 1741	
Ds	1747	T 1747	
RESULT 3	AK082014	LOCUS	
DEFINITION	AK082014	2034 bp mRNA linear	HTC 05-DEC-2002
ACCESSION	AK082014	Mus musculus 16 days embryo head cDNA, RIKEN full-length enriched library, clone: C130094P07 product: hypothetical Glycosyl transferase, family 2 containing protein, full insert sequence.	
VERSION	AK082014.1	GI:26100345	
KEYWORDS	HTC; CAP trapper.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota: Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;		
AUTHORS	Carninci, P. and Hayashizaki, Y.		

TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
99279253  
10349636  
2  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
20499374  
11042159  
3  
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kaishiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Wataniki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
20530913  
11076861  
4  
Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadoya, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, R., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehi, P., Lewis, S., Matsuo, Y., Nikaide, I., Pesole, G., Quackenbush, J., Schriml, L. M., Stauble, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whitaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S. and Hayashizaki, Y.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409 (6821), 685-690 (2001)  
21085660  
11217851  
5  
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
6 (bases 1 to 2034)  
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohashi, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.  
Direct Submission

TITLE  
JOURNAL

Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,















```

Qy 1740 GT 1741
Db 1606 GT 1607

RESULT 7
BE677813
LOCUS
DEFINITION
7f63d08.x1 Soares NSF F8.9W OT PA P S1 Homo sapiens cDNA clone
IMAGE:3299343 3' similar to TR:008832 008832 POLYPEPTIDE GALNAC
TRANSFERASE-T4. ;contains element MER22 repetitive element ;, mRNA
sequence.
ACCESSTON
BE677813
VERSION
BB677813.1 GI:10038280
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 676)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgaps@mail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 477.
Location/Qualifiers
1..676
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3299343"
/lab_host="DH10B"
/clone_lib="Soares NSF F8.9W OT PA P S1"
/notes="Organ: pooled; Vector: pT7T30-Pac (Pharmacia) with
a modified polylinker; Site 1: Not 1; Site 2: Eco RI;
Equal amounts of plasmid DNA from five normalized
libraries were mixed, and ss circles were made in vitro.
Following HAP purification, this DNA was used as tracer in
a subtractive hybridization reaction. The driver was
PCR-amplified cDNAs from pools of 5,000 clones made from
the same 5 libraries. The pools consisted of the following
libraries and clones: Soares NBHSF pool 1:
309384-310919, 323208-325895 Soares Nb2HP pool 1:
145032-147335, 147720-148103, 148872-149255, 15002 -
150407, 151176-152327 Soares Nb2HF8-9W pool 1:
758280-760583, 772104-774407 Soares NbHPA pool 1:
304776-306311, 320136-322823, 326280-326663 Soares NBHOT
pool 1: 723720-726407, 739080-740999 Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 139 a 179 c 221 g 137 t
ORIGIN

Query Match 37.8%; Score 660.8; DB 10; Length 676;
Best Local Similarity 99.0%; Pred. No. 8.4e-138;
Matches 665; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 198 GCGGCGCGCGGTCGCGGCGAAGCGCGTGGCGCGCGCGGCGAGCGGTGGCGTGCAGCT 257
Db 1 GCGGCGCGCGGTCGCGGCGAAGCGCGTGGCGCGCGCGGCGAGCGGTGGCGTGCAGCT 60

Qy 258 GCAGGCGCGAGAGTCGCGTGCAGGAGGAGCGTGCAGTGCACCGAGTTAACATCTA 317
Db 61 GCAGGCGCGAGAGTCGCGTGCAGGAGGAGCGTGCAGTGCACCGAGTTAACATCTA 120

Qy 318 CCTCAGGACCGCATCTCACTGCACCGCGCGTGCAGGCGTGCAGTGCAGTGCAGTGCAG 377
Db 121 CCTCAGGACCGCATCTCACTGCACCGCGCGTGCAGGCGTGCAGTGCAGTGCAGTGCAG 180

378 AGAAGAAGAAATATGATTATGATAATTTGCCAGGACATCTGTTATCATAGCAATTTATAA 437
181 AGAAGAAGAAATATGATTATGATAATTTGCCAGGACATCTGTTATCATAGCAATTTATAA 240

438 TGAAGCCTGGTCAACTCTCTTCGGACAGTTTACAGTGTCTTGGAGACATCCCGGATAT 497
241 TGAAGCCTGGTCAACTCTCTTCGGACAGTTTACAGTGTCTTGGAGACATCCCGGATAT 300

498 CTGCTAGAGAAGAGTGTATCTCTTACAGTGTACAGTGTAGAGAGCAGCCTGAAGGAGCG 557
301 CTGCTAGAGAAGAGTGTATCTCTTACAGTGTACAGTGTAGAGAGCAGCCTGAAGGAGCG 360

558 CTTGGCCAATAGACTTTTCGGGACTGCCCAAGGTGCGCTGATCCGCGCCCAACAGAGAGA 617
361 CTTGGCCAATAGACTTTTCGGGACTGCCCAAGGTGCGCTGATCCGCGCCCAACAGAGAGA 420

618 GGGCCTGGTGCAGAGCCCGGCTGCTGGGGGGCTGCTGGGGCGAGGGCGATGTTCTGACCTT 677
421 GGGCCTGGTGCAGAGCCCGGCTGCTGGGGGGCTGCTGGGGCGAGGGCGATGTTCTGACCTT 480

678 CTGGAAGTGTCACTGTGAGTGCACCAAGGCTGCTGGAGCCGCTGCTGCAGAGGATCCA 737
481 CTGGAAGTGTCACTGTGAGTGCACCAAGGCTGCTGGAGCCGCTGCTGCAGAGGATCCA 540

738 TGAAGAGAGAGTTCGGCAGTGTGTCGCCGCTGATTGATGATCGACTTGAACACCTTCGA 797
541 TGAAGAGAGAGTTCGGCAGTGTGTCGCCGCTGATTGATGATCGACTTGAACACCTTCGA 600

798 ATACCTGGGGAACTCCGGGAGAGCCCGAGATCGCGGCTTCGACTGAGAGGTGTGTTTAC 857
601 ATACCTGTGGAACTCCGGGAGAGCCCGAGATCAGCGGTTTCGACTGGATGCTGTTTTCAC 660

858 GTGGCACACAGT 869
661 GTAGCACACAGT 672

RESULT 8
CB165548/c
LOCUS
DEFINITION
BEQ603160969.R1 CSEQFXN34 liver Bos taurus cDNA, mRNA sequence.
CB165548
ACCESSION
CB165548.1 GI:28151673
VERSION
EST.
KEYWORDS
Bos taurus (cow)
SOURCE
Bos taurus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 692)
Adelson, D.L. and Gill, C.A.
AUTHORS
Bovine ESTs (Adelson and Gill)
TITLE
Unpublished
JOURNAL
COMMENT
Contact: David L. Adelson
Animal Breeding and Genetics
Texas A&M University
Animal Science Dept., TAMU-2471, College Station, TX 77843-2471,
USA
Tel: 9798452616
Fax: 9798456970
Email: david.adelson@tamu.edu.

FEATURES
source
1..692
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="liver"
/clone_lib="CSEQFXN34 liver"
/notes="Organ: liver; Vector: pBluescript SK+; Site 1: NotI
; Site 2: EcoRI; sequence 5' of the insert
(5'-NNN...NNNinsert)
GCCAATTGGAGCTCCACCAGCGGTGGCGGCGGCGGCTCGAG. Sequence 3' of

```

the inserts (AAGAAATTCGATATCAAGCTTTATCGATACCGTCCGACCTCGAG,  
normalized Rd 2 library, sequenced 3' with M13R primer."

```
BASE COUNT      149 a   214 c   174 g   155 t
ORIGIN
Query Match      32.4%; Score 566.4; DB 14; Length 692;
Best Local Similarity 89.6%; Pred. No. 1.4e-116;
Matches 620; Conservative 0; Mismatches 71; Indels 1; Gaps 1;
Qy 330 CATCTCAGTACCGCCGCTGCCGAGCGGTGAACCCGCTGTGCAAGAGAGAATA 389
Db 692 CATCTCAGTACCGCCGCTGCCGAGCGGTGAACCCCTCTGTGCAAGAGAGAATA 633
Qy 390 TGATTATGATAATTTGCCAGGACATCTGTATCATAGCATTTTATATGAAGCTCGTC 449
Db 632 TAAGTAGATGAATGTCGCCAGCATCTGTGCTGATAGCTTTTATATGAAGCTCGTC 573
Qy 450 AACTCTCTCTCGGACAGTTTACAGTGTCTTTGAGACATCCCGGATATCTCTGTAGAAGA 509
Db 572 AACTCTCTCTCGGACAGTTTACAGTGTCTCTCGAGACATCCCGACACCTCTGTAGAAGA 513
Qy 510 AGTGATCTCTGTAGATGACTACGTATAGAGACACTGAGAGCGCTTGGCCAATGA 569
Db 512 AGTTATCTCTGTAGATGACTACGTATGAGAGACACTGAGAGCGCTTGGCCACCGA 453
Qy 570 GCTTTCCGGACTGCCAAGGTGCCCTGATCCGCGCCCAACAAGAGAGAGGCGCTGGTGG 629
Db 452 GCTGGCGGGTGTGCCAAGGTGCCCTGATCCGAGCCCAACAAGAGAGAGGCGCTGGTGG 393
Qy 630 AGCCCGGCTGTGGGGGCGTGTGCGGCGAGGGCGATGTTCTGACCTTCTCTGGAAGTCA 689
Db 392 AGCAGGCTGTAGGGGCTTCCGTGGCCNAGGGCGAGTGTGACCTTCTGGAAGTCCA 333
Qy 690 CTGTGAGTGCCACGAAAGGTGGTGGAGCGCTCTGTCAGAGAGATCCATGAAGAGAGTC 749
Db 332 CTGCGAGTGCATGAGGGGTGGTGGAGCCACTGCTGAGAGAGATCCATGAAGAGAGTC 273
Qy 750 GGCAAGTGTGTCGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 809
Db 272 AGCGGTGGTGGCCCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 213
Qy 810 CTCGGGGAGCCCGAGATCGG - CGGTTTCGACTGGAGGCTGGTGTTCACGTGGCACACAG 868
Db 212 TGCTGGGGAGCCCGAGATCGGAGCGTTTGACTGGCGCTGGTGTTCACGTGGCACACGG 153
Qy 869 TTCTGAGAGAGGAGAGATAGCGATGCAATCCCGCTGCTGATGATGATGATGATGATGATG 928
Db 152 TTCCCGAAAGAGAGAGATCGCATGCGGTGCCCGCTCCCGCATCGATGATGATGATGATG 93
Qy 929 TGGCTGGTGGGCTGTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATG 988
Db 92 TGGCTGGTGGGCTGTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATG 33
Qy 989 CAGGAATGGAAGTTTGGGGAGGAGAAACCTC 1020
Db 32 TAGGATGGAAGTTTGGGGTGAGAAACCTC 1
```

```
RESULT 9
LOCUS      AI800923
DEFINITION wgl4f02.x1 Soares NSF P8_9W OT PA P S1 Homo sapiens cDNA clone
IMAGE:2365083 3' similar to TR:008832 008832 POLYPEPTIDE GALINAC
TRANSEBASE-T4. ; mRNA sequence.
ACCESSION  AI800923
VERSION     AI800923.1 GI:5366395
KEYWORDS   EST.
SOURCE      Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 561)
            NC-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
```

# TITLE

## JOURNAL

### COMMENT

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: [cgaps-r@mail.nih.gov](mailto:cgaps-r@mail.nih.gov)  
This clone is available royalty-free through LNL ; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Insert Length: 915 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 389.  
Location/Qualifiers

#### FEATURES

source

1..561  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2365083"  
/lab\_host="DH10B"  
/clone\_lib="Soares NSF P8\_9W OT PA P S1"  
/notes="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with  
a modified polylinker; Site 1: Not 1; Site 2: Eco RI;  
Equal amounts of plasmid DNA from five normalized  
libraries were mixed, and ss circles were made in vitro.  
Following HAP purification, this DNA was used as tracer in  
a subtractive hybridization reaction. The driver was  
PCR-amplified cDNAs from pools of 5,000 clones made from  
the same 5 libraries. The pools consisted of the following  
libraries and clones: Soares NBHSP pool 1:  
309384-330919, 323208-325895 Soares NB2HP pool 1:  
145032-147335, 147720-148103, 148872-149255, 15002 -  
150407, 151176-152327 Soares NB2HF8-9W pool 1:  
758280-760583, 772104-774407 Soares NBHPA pool 1:  
304776-306311, 320136-322823, 326280-326663 Soares NBHOT  
pool 1: 723720-726407, 739080-740999 Subtraction by Bento  
Soares and M. Fatima Bonaldo."

BASE COUNT 121 a 144 c 177 g 119 t

Query Match 31.4%; Score 548.2; DB 9; Length 561;  
Best Local Similarity 98.6%; Pred. No. 1.5e-112;  
Matches 553; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```
Qy 241 GCGTGGCGGTGCGAGTGCAGGCGGAGAGCTCGGCTGCAGAGAGAGAGCGTGGCGCTG 300
Db 1 GCGTGGCGGTGCGAGTGCAGGCGGAGAGCTCGGCTGCAGAGAGAGAGCGTGGCGCTG 60
Qy 301 CACGAGTTAACTTCTACCTCAGGACCGCATCTCACTGACCGCGCTCCGCGAGCGC 360
Db 61 CACGAGTTAACTTCTACCTCAGGACCGCATCTCACTGACCGCGCTCCGCGAGCGC 120
Qy 361 TGGAAACCGCTGTGCAAGAGAGAAATATGATTATGATAATTTGCCAGGACATCTGTT 420
Db 121 TGGAAACCGCTGTGCAAGAGAGAAATATGATTATGATAATTTGCCAGGACATCTGTT 180
Qy 421 ATCATAGCATTTTATATGAAGCTGTGTCAACTCTCTCCGACAGTTTACAGTGTCTT 480
Db 181 ATCATAGCATTTTATATGAAGCTGTGTCAACTCTCTCCGACAGTTTACAGTGTCTT 240
Qy 481 GAGACATCCCGGATATCTCTGCTAGAGAGATGATCTCTTGTAGATGATGATGATGATG 540
Db 241 GAGACATCCCGGATATCTCTGCTAGAGAGATGATCTCTTGTAGATGATGATGATGATG 300
Qy 541 GAGCACCCTGAAGAGCGCTTGGCCAAATAGCTTTTGGGACTGCCCAAGGTGCGCTGATC 600
Db 301 GAGCACCCTGAAGAGCGCTTGGCCAAATAGCTTTTGGGACTGCCCAAGGTGCGCTGATC 360
Qy 601 CGCGCCAAAGAGAGAGAGGCGCTGGTGGAGCCCGGCTGTCTGGGGGCGCTCTGCGCGAGG 660
Db 361 CGCGCCAAAGAGAGAGAGGCGCTGGTGGAGCCCGGCTGTCTGGGGGCGCTCTGCTGCGAGG 420
Qy 661 GCGGATGTTCTGACCTTCTGAGTGTCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 720
Db 421 GCGGATGTTCTGACCTTCTGAGTGTCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 480
```

Qy	721	CTGCTGCAGAGGATCCATGAAGAGGAGTCGGCACTGTGTGCGCGGTGATTGATGTGATC	780
Ds	481	CTGCTGCAGAGGATCCATGAAGAGGAGTCGGCACTGTGTGCGCGGTGATTGATGTGATC	540
Qy	781	GACTGGGAACCTTCGGAATAC	801
Ds	541	GACTGGGACACCTTCGATAC	561
RESULT 10			
LOCUS	AK033494	2607 bp	mRNA linear HTC 05-DEC-2002
DEFINITION	Mus musculus adult male colon cDNA, RIKEN full-length enriched library, clone:9030401N20 product:UDP-N-acetyl-alpha-D-galactosamin e:polypeptide N-acetyl:galactosaminyltransferase 4, full insert sequence.		
ACCESSION	AK033494		
VERSION	AK033494.1	GI:26329156	
KEYWORDS	HTC; CAP trapper.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1 Carninci, P. and Hayashizaki, Y.		
AUTHORS	High-efficiency full-length cDNA cloning		
TITLE	Meth. Enzymol. 303, 19-44 (1999)		
JOURNAL	99279253		
MEDLINE	10349636		
PUBMED	2		
REFERENCE	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.		
AUTHORS	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes		
TITLE	Genome Res. 10 (10), 1617-1630 (2000)		
JOURNAL	20499374		
MEDLINE	11042159		
PUBMED	3		
REFERENCE	Shibata, K., Itoh, M., Aizawa, K., Nagasaka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kiteunai, T., Tashiro, H., Itoh, M., Sui, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, F., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.		
AUTHORS	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer		
TITLE	Genome Res. 10 (11), 1757-1771 (2000)		
JOURNAL	20530913		
MEDLINE	11076861		
PUBMED	4		
REFERENCE	Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadoya, K., Matsuura, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pebole, G., Quackenbush, J., Schiraldi, L. M., Staubli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, J., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyo-oka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilming, L., Wyshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S. and Hayashizaki, Y.		
AUTHORS	Functional annotation of a full-length mouse cDNA collection		
TITLE	Nature 409 (6821), 685-690 (2001)		
JOURNAL	21085660		
MEDLINE			

11217851

PUBMED  
REFERENCE  
AUTHORS

TITLE

JOURNAL

REFERENCE  
AUTHORS

The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

6 (bases 1 to 2607)

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

Direct Submission

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN.

Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.

URL: http://genome.gsc.riken.go.jp/

URL: http://fantom.gsc.riken.go.jp/

Location/Qualifiers

1. 2607

/organism="Mus musculus"

/mol\_type="mRNA"

/db\_xref="FANTOM DB:9030401N20"

/db\_xref="taxon:10090"

/clone="9030401N20"

/sex="male"

/tissue type="colon"

/clone\_lib="RIKEN full-length enriched mouse cDNA library"

/dev\_stage="adult"

246. 1982

/notes="unnamed protein product;

UDP-N-acetyl-alpha-D-galactosamine:polyptide

N-acetylgalactosaminyltransferase 4 (MGD|MGI:894692,

GB|NM\_015737, evidence: BLASTN, 100%, match=1737)

putative"

/codon\_start=1

/protein\_id="BAC28317.1"

/db\_xref="GI:26329157"

/translation="MAVRVTWAGKSCLLALLTLAVLVEFSVSLYASPCAGAREL

GPRLPDLTREDLSQPLIYKPPADSHALGNGWGRASKQLNGELGKQBELLEIYAI

NYLSVRLSHRHIEDKRYEAKKFKYRSPTTSVIIAFYNEAWSTLLRTHSVLE

TSFVALLKEIILVDSDRIYLAQLEYISLNRVLRITNRKGLRGLIGATFA

TGDVLTFLDCHCECTNGWLEPILERSDEITAVCPVIDTDMNTFEFYMTGPEMIG

GDWRITFQWHSVPKHERDRTRSIDPISPTMAGLGFAVSKVFOYLGTYDTGMEVM

GGENLELSFRVQCCKGLIHCPSHVGHVFPKAPYARNFLONTARAAEVMWDEYKE

HFYNRPAPKAYGDLSEKRLKCKSPDWYLNKVSFNHVDNEDRPGWGAIRS

MGISSCLDYNAPDNNPTGALSLFCGCHGGNQFFEYTSNKIEIRNSVTELCAEVPQ

QKDYGMQNCQKGLFPPVNIWHFKEDGTIFHPHTRCLRLCSAVRTAEGRSPVHMKTCD

ALDKNGLWRFKK"

2578. 2583

/note="putative"

2607

/note="putative"

polyA\_signal

polyA\_site

BASE COUNT 677 a 628 c 660 g 642 t

## ORIGIN

Query Match	29.8%	Score 520.8	DB 11	Length 2607
Best Local Similarity	63.1%	Pred. No. 4.1e-106		
Matches 820	Conservative 0	Mismatches 477	Indels 3	Gaps 1
Qy	201	GCGCCCGTGC	CGCGGAAC	CGCTGGCGCGCGCGCGCGAGCGCGTGCAGCTGCA 260
Db	443	GC	CCCCCTGCAGATTC	CCCATGCTCTTGGGAATGGGGAGAGCAAGCAATCCAGCTCAA 502
Qy	261	GGCGAGAGCTG	CGGCTGCAGGAGGAGCGTGC	CGGCTGCACAGATTAACTATCACT 320
Db	503	CGAGGGGAACT	GAAGCAGCAGAA	GAAGAACTTATTGAGAGATATGCCATCAACATTTAACCT 562
Qy	321	CAGCGACCGCATCT	CATGTCACCGCGCTGCCGAGCGCTG	GNAACCGCTGTGCAAGA 380
Db	563	CAGTGA	CAGGATCTCCCTGCA	CCGCCACATAGAGGATAAAGAATGTATGAGTGTAAAGC 622
Qy	381	GAAGAAATATG	ATTATGATTAATTTGCC	CAGGACATCTGTTATCATAGCATTTTATAATGA 440
Db	623	CAAGAAGTTCC	ACTACAGTTCGCTCCCA	CCACCTCCGTTATCATCGCTTCTATAACGA 682
Qy	441	AGCTGTGTA	ACTCTCTTTCGGA	CAGTTTACAGTGTCTTGTAGACATCCCGGGAATCTCT 500
Db	683	AGCTGTGTC	CACTCTGCTTCG	GACCAATTCACAGTGTTTTAAACACTTCTCCAGCTGTGCT 742
Qy	501	GCTAGAGA	AGTGTATCTTGT	PAGATGACTACAGTGATAGAGACACCTGAAGGAGCGCTT 560
Db	743	TTTAAAGG	AGATCATCTTGT	TGGATGACTTGGATGTAGAAATTTATTGAAGCGCCAACT 802
Qy	561	GGCCAA	TGAGCTTTCCG	GACTGCCCAAGGTGCGCTGATCCGCCCAACAGAGAGAGG 620
Db	803	TGAAACTT	TACATCAGCA	ACCTGGAGAGATTCGCTTGATTAGAACCAATAAGAGAGAGG 862
Qy	621	CCTGGTGG	GAGCCGGCTGTG	GGGGCGCTCTGCGCGAGGGCGCATGTTCTGACCTTCCT 680
Db	863	GCTGGTCC	GGGCCCGCTCTAA	TGTGGGCCACTTTTGGCACCGGGGATGTCTCATCTTCCT 922
Qy	681	GGACTGTCA	CTGTGAGTGCCA	CAGAAGGTGTGCTGGAGCGCTGCTGCAGAGGATCCATGA 740
Db	923	GGATTGTCA	CTGTGAGTGTA	CACTGGTGTGGTGGAAACACTTTTGGAGGATCAGTAG 982
Qy	741	AGAGGAGT	CGGCACTGTGTG	CCCGGTGATTTGATGTGATCGACTGGGAACACTTCGAATA 800
Db	983	AGACGAA	ACCGCAATCG	TATGTCCTGTCAATCGACACTATTGACTTGGAACTTTTTCGAAT 1042
Qy	801	CCTGGGAA	CACTCCGGGAG	CCCCAGATCGCGGTTTCGACTGGAGGCTGGTGTTCAGTG 860
Db	1043	CTACATG	CAGACCGGGAG	CCCATGTTGGTGGGTTTGACTTGGCGTTTGACCTTCAGTG 1102
Qy	861	GCACACAG	TTCTTGAGAGG	GAGGATACCGGATGCAATCCCGCTCGATGTCAATCAGGTC 920
Db	1103	GCATTCGT	CCCCCAACATGA	AGGACAGCGCAAGCTCAAGAAATGACCCCATCAGTC 1162
Qy	921	TCCAA	CAATGCTGGTGG	CTGTTGCTGTGAGTAAGNAATATTTTGAATATCTGGGTC 980
Db	1163	CCCCAC	CAATGCGAGG	ACTGTTGTGTGTCAGCAAGAAGTATTTTTCAGTACCTTGGGCAC 1222
Qy	981	TTATGAT	ACAGAAATGGA	AGTTTGGGAGGAGAAAACTCGAAATTTTCTTTTAGGATCTG 1040
Db	1223	CTATG	CACATGGCATGG	NAAGTCTGGGAGGCGAAAACTTGTAGTTGCTTTTTCAGGTTTG 1282
Qy	1041	GCAGTGTG	TGGTTCGGA	AAACACACCAATGTTTCCCATGTTGGCCATGTTTTCCCAA 1100
Db	1283	GCAGTGTG	CGGCGCAAGCT	TGGAGATCCACCCATGCTCGCACGTGGGCCCACTGTTTCCCAA 1342
Qy	1101	GCAAGCT	CCCTACTCCG	CAACAGGCTCTGGCCAAACAGTCTCGTCGACGTGAAGTATG 1160
Db	1343	CGCGG	CCCTTATGCT	CGCCCACTTCCTACAGAAATACCGCGGGCGCCGAAGTGTG 1402
Qy	1161	GATGGAT	GAAATTTAA	AGAGCTCTACTACCATCGCAACCCCGCTGCCGCTTTGGAACTTT 1220
Db	1403	GATGGAT	GAATACAA	AGGATTTCTATAACCGGAAACCCCGCTGCGAAAGAGGCTTA 1462

Qy	1221	TGGGATGTCACAGAGGGAAGCAGCTCCGGGACAGACTCCAGCTGTAAAGACTTCAAGTG	1280
Db	1463	TGGCGATCTTTTCTGAAGGAAGTACTTAAGGGGAACGGCTGAATGTAAAGACTTTCGACTG	1522
Qy	1281	GTTCCTTGGACACTGTGTATCCAGAACTGCAATGTGCCTTGAGGACACGGCGCTGGCTTCTTCGG	1340
Db	1523	GTATCTGAAAAATGTGTTCTCTTAACCTTACACGCTCCCGAGGATAGACCTGGCTGGCACGG	1582
Qy	1341	GATGCTCCAGAACAAAGGACTTAACAGACTACTGCTTTGACTATAACCCCTCCGATGAAAA	1400
Db	1583	AGCCATCCGCAGCATGGGCATCTCTCCGAGTGTCTAGATTATAACCGCTCCCGACACAA	1642
Qy	1401	CCAGATTGTGGGACACCAAGCTCATCTGTACCTCTGTCACTGGGATGGGCCAGCAATCAGTT	1460
Db	1643	CCCTACGGGGGCTAAATC---TTTCGTTGTTTGGATGCCCGTCAAGGAGGCATCAGTT	1699
Qy	1461	TTTCGAGTACACGCTCCAGAAAGAAATACGCTATAACACC	1500
Db	1700	CTTTGGAATATCTTCAACAAAGAAATACAGATTCATTC	1739

## RESULT 11

LOCUS	BG167520	967 bp	mRNA	linear	EST 06-FEB-2001							
DEFINITION	603242728f1 NIH_MGC_89 Homo sapiens cDNA clone IMAGE:4452799 5', mRNA sequence.											
ACCESSION	BG167520											
VERSION	BG167520.1	GI:12674149										
KEYWORDS	EST.											
SOURCE	Homo sapiens (human)											
ORGANISM	Homo sapiens											
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.											
REFERENCE	1 (bases 1 to 967)											
AUTHORS	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .											
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)											
JOURNAL	Unpublished											
COMMENT	Contact: Robert Strausberg, Ph.D.											

cDNA Library Preparation: Life technologies, inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLAM10242 row: C column: 08  
 High quality sequence spot: 664.

FEATURES source

```

source      1. .967
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:4452799"
            /tissue_type="hypernephroma, cell line"
            /lab_host="DH10B (phage-resistant)"
            /clone_lib="NIH MGC 89"
            /note="Organ: kidney; Vector: pCMV-SPORT6; Site 1: NotI;
            Site 2: SalI; Cloned unidirectionally; oligo-dT-primed.
            Average insert size 1.3 kb. Library enriched for
            full-length clones and constructed by Life Technologies.
            Note: this is a NIH MGC Library."
            276 a    240 c    229 g    222 t
BASE COUNT
          CCGTCN

```

Query Match	29.7%	Score 519;	DB 10;	Length 967;
Best Local Similarity	100.0%;	Prod. No. 6.9e-106;		
Matches 519;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1228	GTGACAGAGGAGCAGCTCCGGGCAAGCTCCAGTGTAAAGACTTCAAGTGGTCTT	1287	
DB	1	GTGACAGAGGAGCAGCTCCGGGCAAGCTCCAGTGTAAAGACTTCAAGTGGTCTT	60	





Db	588	GAGAGGGCTAGTCGAGCCCGCTACTGGGAGCCTCTGCTGCGAGGGGCCGAGTGTCTGA	644
Qy	674	CTTCTCTGACCTGTCTACCTGTGAGTGCCCAAGAGGGTGGCTGGAGCCGCTGCTGCAGAGGA	733
Db	648	CCTTCTAGACTGTCTACCTGTGAATCTCATGAGGGGTGTTGGAGCCTCTGCTGCAGAGGA	707
Qy	734	TCACAGAGGAGGAG 747	
Db	708	TCACAGAGGAGGAG 721	
RESULT 14			
LOCUS	AI863865	511 bp	mRNA linear EST 21-DEC-1999
DEFINITION	wJ55b07.x1 NCI CGAP Lul19 Homo sapiens cDNA clone IMAGE:2406709 3'		
	similar to TR:008832 008832 POLYPEPTIDE GALNAC TRANSFERASE-T4.		
	;contains element MER22 repetitive element ;, mRNA sequence.		
ACCESSION	AI863865		
VERSION	AI863865.1	GI:528058	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 511)		
AUTHORS	NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .		
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),		
JOURNAL	Unpublished		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:cgapbs@mail.nih.gov">cgapbs@mail.nih.gov</a> Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: M. Bento Soares, Ph.D. CDNA Library Arrayed by: Greg Lemmon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://www.bio.llnl.gov/bbbrp/image/image.html">www.bio.llnl.gov/bbbrp/image/image.html</a> Insert Length: 1789 Std Error: 0.00 Seq primer: -40UP from Gibco High quality sequence stop: 453.		
FEATURES			
	source		
	1..511		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone=IMAGE:2406709"		
	/tissue_type="squamous cell carcinoma, poorly		
	differentiated (4 pooled tumors, including primary and		
	metastatic)"		
	/dev_stage="adult"		
	/lab_host="PH103 (phage-resistant)"		
	/clone_lib="NCI CGAP Lul19"		
	/notes="Organ: lung; Vector: pTT3D-Pac (Pharmacia) with a		
	modified polylinker; 1st strand cDNA was prepared from		
	pooled lung tumor tissue, and was then primed with a Not I		
	- oligo(dT) primer. Double-stranded cDNA was ligated to		
	Eco RI adaptors (Pharmacia), digested with Not I and		
	cloned into the Not I and Eco RI sites of the modified		
	pTT3 vector. Library went through one round of		
	normalization. Library constructed by Bento Soares and M.		
	Fatima Bonaldo."		
BASE COUNT	108 a 137 c 167 g 98 t	1 others	
ORIGIN			
Query Match	27.5%;	Score 480.6;	DB 9; Length 511;
Best Local Similarity	98.6%;	Pred. No. 2.3e-97;	
Matches	505; Conservative	0; Mismatches	5; Indels 2; Gaps 2
Qy	198	GCGGCGCGCGTGTCCCGCGAAGCGCTGGCGCGCGGGGCGAGCGGTGCGGCTCGAGCT	257
Db	1	GGGGCGCGCGTGTCCCGCGAAGCGCTGGCGCGCGGGGCGAGCGGTGCGGCTCGAGCT	60



258	Qy	GCAGGGCGAGAGCTGCGGCTGCAGGAGGAGAGCTGCGGCTGCACAGATTAAACATCTA	317
61	Db	GCAGGGCGAGAGCTGCGGCTGCAGGAGGAGAGCTGCGGCTGCACAGATTAAACATCTA	120
318	Qy	CCTCAGGACCGCATCTCACTGCACCCGCGCTCCCGAGCGCTGGAACCCGCTGTGC	377
121	Db	CCTCAGGACCGCATCTCACTGCACCCGCGCTCCCGAGCGCTGGAACCCGCTGTGC	180
378	Qy	AGAGAAGAAATATGATTATGATAATTTGCCAGGACATCTGTTATCATAGCATTTTATAA	437
181	Db	AGAGAAGAAATATGATTATGATAATTTGCCAGGACATCTGTTATCATAGCATTTTATAA	240
438	Qy	TGAAGCTGTGCTCAACTCTCTTCGGACAGTTTACAGTGTCTTTAGAGACATCCCGGATAT	497
241	Db	TGAAGCTGTGCTCAACTCTCTTCGGACAGTTTACAGTGTCTTTAGAGACATCCCGGATAT	300
498	Qy	CCTGCTAGAAGAAGTGATCCTTTGATGATGACTACAGTGATAGAGAGACCTGAAGGAGCG	557
301	Db	CCTGCTAGAAGAAGTGATCCTTTGATGATGACTACAGTGATAGAGAGACCTGAAGGAGCG	360
558	Qy	CTTGCCCAATGAGCTTTTCGGGACTGCCCCAAGGTGGCTGATCCGCGCCAAACAAGAGAGA	617
361	Db	CTTGCCCAATGAGCTTTTCGGGACTGCCCCAAGGTGGCTGATCCGCGCCAAACAAGAGAGA	420
618	Qy	GGG - CTTGTGTGCAGCCCGCTGCTGGGGGCTCTGCGGCGAGGGGCGATGTTCTGACCT	676
421	Db	GGGCTTGTGTGCAGCCCGCTGCTGGGGGCTCTGCGGCGAGGGGCGATGTTCTGACCT	480
677	Qy	TCCTGGACTGTCACTGTGAGTGCACGAAGGG	708
481	Db	T-CTGGACTGTCACTGTGAGTGCAGGAAGGG	511

RESULT	15
AA429394	
LOCUS	
DEFINITION	
	AA429394          496 bp     mrna        linear    EST 16-OCT-1997
	zw32hl2.s1 Soares ovary tumor NHOBT Homo sapiens cDNA clone
	IMAGE:771047 3, similar to TR:E209711 E209711
	UDP-GALNAc:POLYPEPTIDE N-ACETYL GALACTOSAMINYLTRANSFERAS. ; , mRNA
	sequence.

[illegible]

CONTACT: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Possible reversed clone: similarity on wrong strand  
Possible reversed clone: polyt not found  
Seq primer: -41ml3 fwd. ET from Amersham  
High quality sequence stop: 310.

```
FEATURES
source
    1. .496
    Location/Qualifiers
    high quality sequence stop: 310.
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="GDB:5980921"
    /db_xref="taxon:9606"
```

```

/clone="IMAGE:77104"
/sex="Female"
/tissue_type="ovarian tumor"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares ovary tumor NbHOT"
/note="Organ: ovary; Vector: pT713D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
TGTTCAACTCTCAAGTCGAGCGCGCGGTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT713 vector
(Pharmacia). Library constructed by Bento Soares and
M.Fátima Ronaldo."
107 a 157 g 100 t
BASE COUNT
107 a 157 g 100 t

```

BASE COUNT	107 a	132 c	157 g	100 t	ORIGIN
Query Match	25.8%	Score 450.8	DB 9	Length 496	
Best Local Similarity	96.8%	Pred. No. 1.1e-90			
Matches 482; Conservative	0	Mismatches 12	Indels 4	Gaps 2	
QY	203	CGCGGTGCGGGCGAACGGCTGGCGCGCGCGGGCGAGAGCGCGTGCAGCTGCAGG	262		
Db	2	CGCGCGCGGTGCCGGAACGGCTGGCGGG--GGAGCGAGCGCGTGCAGCTGCAGTGCAGG	58		
QY	263	GCAGGAGCTGCGCGTGCAGGAG-GAGAGCGTGGCTGCACAGATTTAACATCTACCTC	321		
Db	59	GCAGGAGCTGCGGCTGCAGGAGAGAGCGTGGACTGCACAGATTTAATCTACCTC	118		
QY	322	AGCGACCGCATCTCACTGCACCGCGCGCTGCCGAGCGCTGGAAACCGCTGTGCAAAAGAG	381		
Db	119	AGCGACCGCATCTCACTGCACCGCGCGCTGCCGCTGGAAACCGCTGTGCAAAAGAG	178		
QY	382	AAGAAATATGATTATCATAAATTTGCCAGGACATCTGTTATCATAGCATTTTATAATGAA	441		
Db	179	AAGAAATATGATTATGATTAATTTGCCAGGACATCTGTTATCATAGCATTTTATAATGAA	238		
QY	442	GCCTGTCAACTCTCCTTCGGACAGTTTACAGTGTCTCTTGAGACATCCCGGATATCCTG	501		
Db	239	GCCTGTCAACTCTCCTTCGGACAGTTTACAGTGTCTCTTGAGACATCCCGGATATCCTG	298		
QY	502	CTAGAGAAAGTGAATCCTTGTAGATGACTACAGTGATAGAGACACCTGAGGAGCGCTTG	561		
Db	299	CTAGAAAGAGTGAATCCTTGTAGATGACTACAGTGATAGAGACACCTGAAAGGAGCGCTTG	358		
QY	562	GCCAAATGAGCTTTTCGGGACTGCCCAAGTGGCGCTGATCCGCGCCCAACAAGAGAGAGGGC	621		
Db	359	GCCAAATGAGCTTTTCGGGACTGCCCAAGTGGCGCTGATCCGCGCCCAACAAGAGAGAGGGC	418		
QY	622	CTGGTGCAGACCGCGCTGCTGGGGCGTCTGGCGGAGGGGCGATGTTGACCTTCCTG	681		
Db	419	CTGGTGCAGACCGCGCTCTGGGGCGTCTGGCGGGAAGGGCGATGTTGACCTTCCTG	478		
QY	682	GACTGTCACTGTGAGTGC	699		
Db	479	GACTGTCACTGTGAGTGC	496		

Search completed: December 7, 2003, 21:17:09  
Job time : 3640.14 secs

**THIS PAGE BLANK (USPTO)**